A Solution to the Cosmological Problem of Relativity Theory

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy
in the Department of Physics and Engineering Physics
University of Saskatchewan
Saskatoon

By
Daryl Janzen

Permission to Use

Head of the Department of Physics and Engineering Physics
Physics Building
116 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5E2

Abstract

After nearly a century of scientific investigation, the standard cosmological theory continues to have many unexplained problems, which invariably amount to one troubling statement: we know of no good reason for the Universe to appear just as it does, which is described extremely well by the flat ΛCDM cosmological model. Therefore, the problem is not that the physical model is at all incompatible with observation, but that, as our empirical results have been increasingly constrained, it has also become increasingly obvious that the Universe does not meet our prior expectations; e.g., the evidence suggests that the Universe began from a singularity of the theory that is used to describe it, and with *space* expanding thereafter in *cosmic time*, even though relativity theory is thought to imply that no such objective foliation of the spacetime continuum should reasonably exist. Furthermore, the expanding Universe is well-described as being flat, isotropic, and homogeneous, even though its shape and expansion rate are everywhere supposed to be the products of local energy-content—and the necessary prior uniform distribution, of just the right amount of matter for all three of these conditions to be met, could not have been causally determined to begin with. And finally, the empirically constrained density parameters now indicate that all of the matter that we directly observe should make up only four percent of the total, so that the dominant forms of energy in the Universe should be dark energy in the form of a cosmological constant, Λ, and cold dark matter (CDM).

The most common ways of attacking these problems have been: to apply modifications to the basic physical model, e.g. as in the inflation and quintessence theories which strive to resolve the horizon, flatness, and cosmological constant problems; to use particle physics techniques in order to formulate the description of dark matter candidates that might fit with observations; and, in the case of the Big Bang singularity, to appeal to the need for a quantum theory of gravity.

This thesis takes a very different approach to the problem, in hypothesising that, because our physical model really does appear to do a very good job of *describing* the observed cosmic expansion rate, and all the data indicate that our Universe might well expand precisely according to the flat ΛCDM scale-factor, it may not be the model, but our basic expectations that need to be modified in order to derive a physical theory that stands in reasonable agreement with the empirical results; i.e., that it may actually be that we need to re-examine, and rationally modify our expectations of what should theoretically be, so that we might derive a theory to *explain* the empirical results of cosmology, which would be based solely on reasonably acceptable first principles.

Therefore, a self-consistent theory is constructed here, upon re-consideration of the cosmological foundations of relativity theory, which eventually does afford an explanation of the cosmological problem, as it provides good reason to actually *expect* observations in the fundamental rest-frame to be described precisely by the flat ΛCDM scale-factor which has been empirically constrained.

Acknowledgements

I would like to thank Rainer Dick for advising me to do what interests me, and for allowing me the freedom to do just that. Thanks also to my advisory committee, especially Artur Sowa and Kaori Tanaka, for some very helpful suggestions that greatly improved readability of the final draft. Mostly, though, I want to thank my family, who supported me in every capacity of life while I worked on this thesis.

For Kim, Amalia, and Jonas,
who make everything possible.

Contents

List of Figures

Glossary of Common Terms

BAO	Baryon acoustic oscillations
CBR	Cosmic background radiation
ΛCDM	The two-component cosmological standard model, involving Λ and cold dark matter
CMB	Cosmic microwave background, the present CBR
ESSENCE	Equation of State: SupErNovae trace Cosmic Expansion
FLRW	Friedman-Lemaître-Robertson-Walker
Λ	The cosmological constant
PS_3K	A cosmological theory proposed in three 1967 articles: one, by American physicists, V. Petrosian, E. Salpeter, and P. Szekeres; the other two, separately, by Soviet physicists, J. Shklovsky and N. Kardashev.
RW	Robertson-Walker
SdS	Schwarzschild-de Sitter
SDSS	Sloan Digital Sky Survey
SN(e) Ia	Type Ia supernova(e)
WMAP	Wilkinson Microwave Anisotropy Probe

Introduction

> One should not pursue goals that are easily achieved. One must develop an instinct for what one can barely achieve through one's greatest efforts.
>
> —Albert Einstein

> Nothing will ever be attempted, if all possible objections must be first overcome.
>
> —'the artist', in Samuel Johnson's *Rasselas*

This thesis began from an idea, that there might be some way to describe not only the current accelerated rate of cosmic expansion, but actually the expansion of the Universe at all times, as a consequence of the basic metrical properties of physical reality. Specifically, the basic idea I had, which actually isn't what the eventual theory comes to describe, was that if our Universe had some essential characteristic which would cause it to naturally expand right from the Big Bang, which would be described by a pure cosmological constant, then at early times, when all the mass in the universe was close together, the mutual attraction would have a slowing effect on the rate of that intrinsic cosmic expansion, which would weaken as the universe did expand, until an inflection point was reached and Λ began to dominate.

The problem with this idea is that it can't exactly be described by Friedman's equations, which are totally dominated by material densities near the big bang, so that no matter how colse the expansion of a Friedman-Lemaître-Robertson-Walker (FLRW) universe might actually correspond to this description—e.g., as in the ΛCDM models—the description given by the standard cosmological model and the one I was thinking of could never be alike close to the Big Bang.[1] This is why the classical FLRW 'big bang' models *require* that singular push to start things off, which might be subsequently helped along, e.g. by an inflationary epoch, which would also smooth things out.

Basically, this dynamical theory (and my idea wasn't much different, except that I preferred to think of Λ, and its role for expansion, as a basic metrical property) requires a

[1] Note, that throughout the thesis, I use capitalisation to distinguish the names of (analytically uncertain) things that actually are, such as the Universe, Truth, and Physical Reality, from the things of physical theory, such as a universe, truth, or physical reality; e.g., in the end, after working out the physical properties of a certain universe, I propose that that is what our Universe actually is, with its big bang describing the Big Bang we believe to have taken place at the beginning of Time—which is something that also wants a more concrete analytical description.

dynamical origin for expansion, and I thought, instead, that there should be a way of describing it purely as a metrical property;[2] so, because I still thought there could be something more to this idea of expansion, I began looking into other options. This seemed reasonable enough to do, being that the empirical evidence does generally favour a pure cosmological constant, which, if it were a fundamental constant of Nature, should reasonably be thought to have an essential connection to the cosmic expansion that appears to be a fundamental property of the Universe.

Now, in addition to the fact that the standard cosmological model is incapable of providing an essential explanation of cosmic expansion—I mean, of the most basic cause of cosmic expansion, which it only describes as the consequence of an initial singularity—there are many other well-known problems with it, as the parameters that have been constrained through empirical modelling are, according to the most objective insights that can be made consistently within the dynamical theory, *completely* unrealistic. To me, all of its problems suggested that there really could be something essentially wrong with the standard model, so that rather than lumping conjectures onto that model in order to smooth out the issues it is known to have, the essential basis of the theory should be carefully examined.

And so I began to investigate the details of the historical development of the theory, with the idea that I should be able to find an argument there for how the standard observational model could pertain to an essentially different physical framework that would support the metrical interpretation of cosmic expansion; as, in the meantime, I had already come across a formal mathematical result which strongly indicated to me that this should be possible.

One of the most significant discoveries for me, which I only made in carrying out this investigation, was that the interpretation of Λ as the essential cause of cosmic expansion is actually the one that Arthur Eddington had always fought desperately for; however, he didn't principally think of this in the way of a requisite explanation of cosmic expansion, as he had already begun cultivating the idea in the early 1920s, but actually as the most natural consequence of general relativity theory, as he considered, long before Hubble's confirmation of the cosmic expansion, that the theory could not possibly do without Λ as a fundamental metrical constant [1], even to the extent that he considered the prospect of dropping the cosmological constant as being tantamount to knocking the bottom out of space [2].

To briefly sum up the first chapter of the thesis, then:—it is a detailed examination of the development and current status of modern cosmology, with special attention paid to the works which regarded Λ in one way or another, and an eye to the development of a new theory that would be completely consistent with Eddington's interpretation of cosmic expansion.

[2] I should like to be absolutely clear, that I've never subscribed to the opinion that formal equivalence of the physical descriptions of two essentially different things means as much as if they Truly were identical, and have always felt that there is epistemological merit in trying to understand what things Really are. The contrary position is, as I understand it, one that was maintained by some of Johannes Kepler and Galileo Galilei's strongest adversaries, who eventually conceded the superiority of Kepler's heliocentric system only because it provided a more accurate description of the phenomena. But Kepler and Galileo did not have the luxury of those results when they fought their battles: they led the Copernican revolution in their day because a Truly heliocentric system would give rise more naturally to the phenomena. Of course, the fact that the empirically more accurate Keplerian theory played an essential role in Isaac Newton's eventual discovery of the far deeper theory of Universal gravitation, which allowed him to further refine the description of planetary orbits, is a clear demonstration of the fact that it really is worthwhile to search for essential Truth.

Now, the initial mathematical result referred to above, which really paved the way for this and the rest of the analysis, came from studying the solution which describes the existence of a Schwarzschild mass in de Sitter space,

$$ds^2 = -\frac{r}{r^3 - {r_0}^3 - (r - r_0)}dr^2 + \frac{r^3 - {r_0}^3 - (r - r_0)}{r}dt^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2), \qquad (1)$$

which is reproduced here in the dimensionless coordinate system that is extensively analysed in Chapter 4. The Schwarzschild-de Sitter (SdS) line-element really has two distinct forms: it describes the local gravitational field outside a non-rotating, uncharged spherical mass; but it is also a cosmological solution, in the case ${r_0}^2 > 4/3$, when $r > 0$ is timelike (in which case the explicit form given in Eq. (1) directly applies). The important discovery, was that according to an observer who 'naturally evolves' only according to the gravitational field described by this solution, 'space' is perfectly flat, and r increases with proper time, τ, as

$$r(\tau) = (2M)^{1/3}\sinh^{2/3}(3\tau/2)\,, \qquad (2)$$

where $2M \equiv r_0 - {r_0}^3$ is a constant—which is exactly the form of the scale-factor in the flat ΛCDM model.

The problem, however, is that it is not at all obvious how these coincidences could be related to the standard cosmological model, primarily because the slices of constant r are not synchronous in this frame, which is tilted with the spacelike singularity at $r = 0$ always at one extent of synchronous hypersurfaces, with the other spacelike direction extending towards $r = +\infty$.[3] But then there seemed to be a deeper problem with this picture, because the solution actually describes *all* particles at constant spatial coordinates as being 'naturally at-rest'—so it does not make sense that the bundle of worldlines of all such particles would be *synchronously* evolving, with particles at *all* values of r, including those 'continuously ejected from the singularity at $r = 0$', comoving as they evolve through spacetime, because the gravitational potential drastically varies throughout those synchronous slices.

In fact, the only thing that would make sense is if these particles would be comoving in the r-direction, along spatial slices that are tilted in this frame, so that they would always be at points of spacetime with equivalent gravitational potential.

But then it becomes important to consider particles that move through space in this coordinate system, which are therefore not 'naturally at-rest'. Would these particles, as they evolve with the whole system, not remain in the same comoving hypersurface as the coherent bundle of geodesics just considered? Or would a particle that moves through space in one direction find itself existing at values of r through which the rest-frame particles had already passed, and another, moving in the opposite direction, come to larger r 'before' everything else?

Again, the only thing that makes sense is that there would be an absolute cosmic time, described so that every particle, no matter how it would move through space, would exist on the same causally coherent hypersurface evolving along increasing r from a common origin at $r = 0$.

Because of the similarity of solutions, this problem then got me thinking about the validity of the standard black hole picture, as I began to wonder whether it should really be

[3] It would be helpful to refer to Fig. 4.3, while reading this and the following few paragraphs.

possible for an astronaut existing inside a black hole to shine two flashlights in 'opposite directions', in such a way that the light from one would reach the final singularity 'before' the other (consider, e.g., the evolution of null lines beyond the horizon in synchronous slices of Eddington-Finkelstein or Kruskal-Szekeres coordinates). Would this not mean that the photons that left the one flashlight would, throughout their cosmic evolution towards the final singularity, exist, in a realistic sense, *later* than the astronaut, while those of the other, which were somehow shone through spacetime 'towards' the past horizon, would exist *earlier*? How should the notion of time that is implied by such speculation actually be justified in a causally coherent way, when r is a timelike dimension, according to which things can even be said to 'happen'? i.e., when, in the four-dimensional spacetime continuum, no matter how we 'extend' the coordination of events, r should never be thought of as a dimension through which things can be said to 'move', but really as the denominator of the verb (if we should remain consistent in our interpretation of the meaning of the description)?

Couldn't it then be instead, that regardless of the frame in which the evolution of world-lines is considered, particles travelling along every null line emanating from some given event at some $r < 2m$, should progress through spacetime, along with any other particle that might be said to have existed at that same r 'simultaneously' with that event, so that they should also be said to exist at all later r, including the final singularity, 'simultaneously' as well? i.e., that the existence of photons and astronauts inside a black hole could be interpreted as a comoving evolution along r, no matter the frame?

Similarly, we should say that there is no way of shining a flashlight in any direction of the Universe so that we could reasonably say that its light 'now' exists, in four spacetime dimensions, five minutes ago, and 'now' six, with timelike separation increasing, as everything also dynamically evolves in *another sense*, 'forward in time'. For every point in those four spacetime dimensions corresponds to an event that should take place in the Universe, so that if we want to describe instead the real motion of a particle *through* something like a spacetime manifold, we need to account consistently—and analytically—for that additional notion of time.

The two facts, that the gravitational field is weak near the horizon in most black hole solutions, and that the horizon is always at $r = 2m$, regardless of the coordinate system used, suggest that there really is something wrong with this general picture, in which particles are thought to continually fall through the horizon and continue on towards $r = 0$ just as though r were another spatial dimension—although, of course, one in which everything is 'trapped' and must evolve uni-directionally—through which a particle's trajectory could be mapped, and the same path might be followed 'some time later' by another particle. For there is really no good reason to suspect, that just beyond the horizon the relativistic description should be any different from any other cosmological solution; and we should never say that a rocket flying randomly through space should accidentally find itself out of sync with the presently evolving cosmic time of a FLRW universe, just because instants of cosmic time are not synchronous in its local coordinate system and it measures proper time at a different rate; nor should we ever think of our Universe as truly evolving in cosmic time according to the standard model, while another universe, just like ours, that 'big banged' a billion years later, is somehow evolving through the same four-dimensional manifold as us (as long as we think of things in terms of black hole physics, in which spacetime is like a four-dimensional arena that test-particles move through), a billion years back in the cosmic time direction;

and we definitely should not think of rockets as being able to fly between the two.

And so the question, as regards the standard picture of black holes, is this: is it really correct, or even consistent, to conclude,—based on the fact that there are coordinate systems in which the trajectories of two particles can be drawn so that both fall towards $r = 0$, from $r > 2m$, at different 'times', while a third particle 'remains' outside,—that any star can have collapsed to a singularity, and more particles should continue to flood in through the horizon to the same ultimate fate, moving through the four spacetime dimensions there, in a causally coherent sense during the course of the Universe's real evolution? The answer, I believe, is no. And the resolution of this problem, which is the same as my original problem of needing a coherent description of cosmic evolution in the cosmological SdS solution, Eq. (1), unfortunately requires a different interpretation of the *meaning* of general relativity theory—although one which is totally consistent with relativistic cosmology, and really explains why the standard model, which is consistent with both interpretations, is the most trivial possibility.

Basically, what I needed to argue, was that Einstein was already wrong when, in developing special relativity theory, he took the principle of relativity, that the laws of physics should have the same description in all frames, along with the relativistic description of simultaneous phenomena, to mean that no objective significance may be attached to any particular coordinate system, or, in another way, that the synchronous evolution of things in any frame always provides a good—ontic—noumenal description of simultaneity.

Instead, what I needed to justify was that in all general relativistic solutions the timelike dimension actually best describes the absolute cosmic evolution of a specific three-dimensional hypersurface of spacetime, which is described just as well in all frames, according to the general covariance of the coordinate transformations—and beyond that, in a manner that is roughly consistent with the common idea of the motion of matter within a black hole's event horizon, that this real evolution occurs as a three-dimensional universe evolves dynamically *through* a four-dimensional manifold, so that spacetime should not be thought of as something that really exists at all, but as the map of noumena that occur throughout that course of evolution, which can be similarly described in the proper coordinates of any reference frame, so that different noumena would be given to occur as simultaneous phenomena in different coordinate systems, but with a greater sense of simultaneity pertaining to the well-defined evolution of the universe, which would generally not be synchronous; i.e., that every event on a spacetime manifold corresponds to a hypersurface of events which truly occurred simultaneously on a different four-dimensional manifold through which that hypersurface really moves in a well-defined way, which 'motion' is the same as the cosmic time of modern cosmology, as well as absolute time according to Newton's definition [3]:

> Absolute, true, and mathematical time, of itself, and from its own nature, flows equably [i.e., uniformly] without relation to anything external, and by another name is called duration: relative, apparent, and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a day, a month, a year.

Apart from a need to sort out how general relativity could be consistently assimilated into such a cosmoogical framework—e.g. because relativity theory is generally supposed to be at-odds with the absolute time of Newtonian physics, as indicated by the difference between the Galilean and Lorentz transformations of spacetime coordinates, or because a description of the warping of spacetime around massive particles should need to be consistently accounted for—there were other, more long-standing issues that needed to be addressed, such as Zeno's paradox, that 'if everything that exists has a place, place too will have a place, and so on *ad infinitum*' [4]; in other words, because by adding yet another timelike dimension to the theory—a real fourth dimension other than the one of spacetime, which was now supposed to correspond to observable reality only at the present hypersurface—the theory now was, philosophically speaking, treading on dangerously thin ice.

However, I did have what I now consider to be an ill-conceived notion of time in spacetime, according to the black hole picture, to guide me; for although the 'extended' coordinates of Schwarzschild spacetime remain four-dimensional coordinate systems, they still do, in the way that their causal description has been interpreted, support such a five-dimensional idea; i.e., because they describe a four-dimensional set, $\{r, t, \theta, \phi\}$, bounded in r $(< 2m)$ and spherically closed, which *evolves* in such a way that particles can be *continually dropped into it.* So although such coordinations of the Schwarzschild metric do not explicitly add another dimension, but only 'extend' these ones, it is, I believe, through mis-interpretation of the causal evolution that is described in the new systems, according to that same aspect of Einsteinian relativity that I needed to argue against, that this fifth dimension is anyhow realised.

At this point, the project had really become one on the philosophy of Time, which is a very old problem, and so the rigorous justification I needed to give for the extra-dimensional type of 'presentist' theory I wished to propose required analysis of the basic developments that were made on both sides of the argument, throughout history. The following is a very brief synoptical account of that:—

Twenty-five hundred years ago, Parmenides of Elea had a thesis: all of space and all of time simply exist as a singular (spherical) block universe, and the perceptions of duration, free will, and change are purely illusory. Actually, that statement is not even correct, because Parmenides thought that space and time couldn't actually exist at all, but that all that exists is substantive matter, which therefore exists as one four-dimensional *plenum*. Parmenides worked out this theory deductively, from a single axiom: What-is is all there Is.

In response to Parmenides, the Atomists, Leucippus and Democritus, proposed a new axiomatic physical principle: What-is-not, which they called the 'void', is every bit as Real as What-is, which is atoms of all shapes and sizes, and the atoms move about in the void, bringing about change as they do so.

But the Atomists' void was so at-odds with the common sense-based notion that there truly is substance everywhere, with air or liquid filling even the tiniest pores of every otherwise solid object, that even Aristotle, who sought the same goal of breaking through the barrier that Parmenides had set up, and thereby reconciling his analytic physical theory with fundamental change in reality, in accordance with the Heraclitean thesis of time as the uniform flux of all things, chose instead to admit the requirement of a supernatural prime mover who would achieve the same end as the Atomists' void—as he did allow for the reality of 'place', but developed his theory so that this 'place' would be continuously full of matter, which *rearranged* itself in time, according to a well-ordered duration that was achieved through the abstract metaphysical prime mover.

Centuries later, Newton developed his theory as one that was very much in line with the principles of ancient atomism, which did not admit the presence of a supernatural being anywhere in physical reality except in the initial set-up of the solar system. According to Newton's theory, things moved through absolute space and time which are ontologically as real as the substantial material that exists in them, as the whole system of physical reality would equably endure, with the same order that Aristotle's prime mover was intended to achieve, which requirement was probably made the most clear after Aristotle, through St. Augustine's rationalisation.

It is my opinion that Einstein, who, e.g. along with Ernst Mach and René Descartes, was more inclined towards the Parmenidean way of thought when it came to the age-old debate on the independent reality of a substantive void, than, say, the Atomists and Newton, actually *developed* or *cultivated* his interpretation of the meaning of relativity theory so that it would be in line with that position, which always inevitably has led through, deductively, to the requirement of a block universe.

This is evidenced, e.g., in 'Relativity and the Problem of Space', the fifth appendix to *Relativity: The Special and General Theory*—e.g., where he argued that [5]

> . . . . Mach, in the nineteenth century, was the only one who thought seriously of an elimination of the concept of space, in that he sought to replace it by the notion of the totality of the instantaneous distances between all material points. . . .
>
> [In Newtonian mechanics], "physical reality", thought of as being independent of the subjects experiencing it, was conceived as consisting, at least in principle, of space and time on one hand, and of permanently existing material points, moving with respect to space and time, on the other. The idea of the independent existence of space and time can be expressed drastically in this way: If matter were to disappear, space and time alone would remain behind (as a kind of stage for physical happening).
>
> . . .
>
> We are now in a position to see how far the transition to the general theory of relativity modifies the concept of space. In accordance with classical mechanics and according to the special theory of relativity, space (space-time) has an existence independent of matter or field. In order to be able to describe at all that which fills up space and is dependent on the coordinates, space-time or the inertial system with its metrical properties must be thought of at once as existing, for otherwise the description of "that which fills up space" would have no meaning.* On the basis of the general theory of relativity, on the other hand, space as opposed to "what fills space", which is dependent on the co-ordinates, has no separate existence. Thus a pure gravitational field might have been described in terms of the g_{ik} (as functions of the coordinates), by solution of the gravitational equations. If we imagine the gravitational field, i.e. the functions g_{ik}, to

*If we consider that which fills space (e.g. the field) to be removed, there still remains the metric space in accordance with

$$ds^2 = {dx_1}^2 + {dx_2}^2 + {dx_3}^2 - {dx_4}^2, \tag{3}$$

which would also determine the inertial behaviour of a test body introduced into it. [Einstein's footnote].

> be removed, there does not remain a space of the type [Eq. (2.1)], but absolutely *nothing*, and also no "topological space". For the functions g_{ik} describe not only the field, but at the same time also the topological and metrical structural properties of the manifold. A space of the type [Eq. (2.1)], judged from the standpoint of the general theory of relativity, is not a space without field, but a special case of the g_{ik} field, for which—for the coordinate system used, which itself has no objective significance—the functions g_{ik} have values that do not depend on the co-ordinates. There is no such thing as an empty space, i.e. a space without field. Space-time does not claim existence on its own, but only as a structural quality of the field.
>
> Thus Descartes was not so far from the truth when he believed he must exclude the existence of an empty space. The notion indeed appears absurd, as long as physical reality is seen exclusively in ponderable bodies. It requires the idea of the field as the representative of reality, in combination with the general principle of relativity, to show the true kernel of Descartes' idea; there exists no space "empty of field".

Therefore, apart from the fact that he really was required by developments of modern physics to admit that there is vacuous space between particles of matter, Einstein really had come back as close as he could to Parmenides' theory of one block universe. In fact, the relativistic description of a block universe was first expounded by Hermann Minkowski, when he invented 'spacetime' explicitly as an absolute four-dimensional world in 1908 [6].

In 1966, rigorous proofs were given by Wim Rietdijk [7] and Hilary Putnam [8] which confirmed that the standard interpretation of special relativity theory due to Einstein, that every inertial observer is objectively as justified in claiming that the simultaneous phenomena they observe in their frames are really simultaneous noumena, really does require a block universe. Actually, this is obvious: if we integrate all possible simultaneities of all possibly existing simultaneous special relativistic observers, even at one particular instant in a specified frame, we find that the four dimensions of Minkowski spacetime must absolutely exist simultaneously, so that we can 'imagine that everywhere and everywhen there is something perceptible', as Minkowski put it [6].

The Rietdijk-Putnam determinacy proofs were arguably strengthened later by Howard Stein's [9, 10] attempt to refute them [11, 12]. For, by arguing that special relativistic spacetime will be indeterminate if, and only if,[4] every event outside any observer's causal past is, in a general or absolute sense corresponding to each observer individually, indeterminate here-now, Stein effectively showed that any such 'light-cone argument' in favour of indeterminacy should fail if it is possible to say that even one event that lies outside our causal past, such as the emission of a photon towards the Earth, from *somewhere* on the Sun's photosphere, *sometime* in the past eight minutes, should have occurred without our exact knowledge of the where and when of it. Thus, e.g., by positing that 'at least one event in the universe shares its present with another event's present', which he considers to be 'the thinnest requirement one could put on becoming', Craig Callender transforms 'Stein's "possibility" theorem into a "no go" theorem for objective becoming in Minkowski spacetime' [11].

[4] The argument only works if spacetime is all that really exists, and no more structure is added to it, as it exists in this indeterminate state.

Stein's theory basically corresponds the description of Minkowski spacetime given by Eddington in *The Nature of the Physical World* [13]. In my opinion, such a radically phenomenalist standpoint, which goes contrary to all reason just because it can, according to the principle of scientific verifiability, is metaphysically useless. Even so, it is worth briefly considering whether Stein's proposal for reconciling becoming with relativity theory should have anything to do with standard physics.

To begin with, we might contrast Stein's idea with the basic justification for the standard black hole picture; for then it should be immediately obvious how inconsistent it would turn out to be, if any physicist would claim that they seriously subscribe to this view rather than the Einsteinian one—as contemporary wisdom has it that black holes presently exist all throughout the universe, even though it is impossible that any event that would have occurred at a timelike radius inside a black hole's event horizon, or even precisely at the horizon, could ever find its way into the past light-cone of an external observer. In fact, according to Kip Thorne [14], the basic result that convinced the global community of theoretical physicsts that black holes should presently exist in the Universe, was David Finkelstein's [15] discovery of a coordinate system for the Schwarzschild metric, that describes the synchronous evolution of an observer who remains outside, along with a particle that passes through the horizon and on to the central singularity, in a finite interval of coordinate time.

However, according to Stein, all points outside the past light-cone of here-now mustn't yet have come to be if relativity theory is to describe anything but a four-dimensional block universe; so it cannot be concluded that any particle anywhere in the entire Universe has ever dynamically reached the event horizon of a black hole, because all such events should necessarily have to fall outside the causal past of every external observer—i.e., there is no way of saying definitely that the proper time of any particle that is taking part in gravitational collapse has ticked past whatever finite interval it should take to get arbitrarily close to the horizon. The interpretation of Finkelstein's result as providing evidence that the process of collapse will already have occurred therefore requires the assumption that the description of synchronous space has objective meaning in that frame—and thus, probably every other, because there is absolutely no justification for assuming that Eddington-Finkelstein synchronicity should be somehow special,—which in turn demands a block universe if one actually cares to be honest, along with Einstein,[5] about the formal implications of one's principal interpretation of the theory.

Another example of the fact that relativists have, in principle—if not rigorously in subsequent application,—commonly sided with Einstein et al. rather than Stein, when interpreting the meaning of relativity, is Roger Penrose's treatment of the relativity of simultaneity in *The Emperor's New Mind* [17], where he presents a similar argument to the one given by Rietdijk and Putnam, noting the same implication for determinism. In fact, Penrose even goes so far as to claim, in an endnote to his statement that 'The "now" according to one observer would not agree with that for another', that 'Some relativity "purists" might prefer to use the observers' light cones, rather than their simultaneous spaces. However, this makes no difference at all to the conclusions.' In fact, it makes no such difference *except* in the extreme case proposed by Stein—which, we must therefore infer, Penrose did not register as

[5] Thus, he admitted that 'To us believing physicists the distinction between past, present, and future has only the significance of a stubborn illusion' [16].

a realistic possibility.

Therefore, although we do commonly think of general relativity as a dynamical theory, which would describe the continually evolving local interaction of matter with a four-dimensional metric space that it dynamically warps, the original interpretation of the meaning of the general relativity of simultaneity, due to Einstein, which is still commonly assumed by physicists, leads deductively through to a block universe description, according to which the supposed dynamism is not fundamental.

As Robert Geroch put it [18], 'There is no dynamics within space-time itself: nothing ever moves therein; nothing happens; nothing changes. . . . one does not think of particles "moving through" space-time, or as 'following along" their world-lines. Rather, particles are just "in" space-time, once and for all, and the world-line represents, all at once, the complete life history of the particle.'[6] Or, as he finally concluded, '[Perhaps the distinctive feature of Einstein's theory of relativity] is the introduction of the idea of space-time—the idea that everything of physical interest in the world, including oneself, is to be set down therein once and for all. In the Galilean view, space-time is a luxury; in relativity, a necessity.'

This direct consequence of the standard Einsteinian interpretation of the theory—which implies that there is absolutely no purpose in undertaking any more physical investigations that might lead us to a better understanding of the Universe, since all of it, and everything we do in our lives, should be strictly determined anyway, all of it substantively existing as a fixed four-dimensional block—is not only the most significant, but also the most commonly overlooked problem of modern physics; and it should be impractical, if not impossible, to try to resolve any of the other problems connected with relativity theory so long as it is inconsistently treated, within the standard framework, as a fundamentally dynamical theory, and thereby used to describe pseudo-dynamical processes that are essentially paradoxical.

But as Stein correctly pointed out, first in [9], in relation to Putnam's argument, but also in [19], in relation to Parmenides', 'An argument that leads to a truly paradoxical conclusion is always open (if it escapes conviction for fallacy) to construction as a *reductio ad absurdum*.' This is precisely what the ancient Atomists did, in order to reconcile a physical theory of motion, according to the (otherwise arguably gratuitous) premiss that space really does have an independent ontological existence. For the case of special relativity theory, it is worthwhile to reproduce Stein's more detailed critique of Putnam's argument as well [9]:

> . . . Putnam's conclusion—that all things, for all time, are real—contradicts the "man on the street's view." What sort of argument is it, then, that contradicts its own premise?
>
> From the strictly deductive point of view, such an argument is a *reductio ad absurdum*, i.e., a refutation of the conjunction of its premises; and if we consider the "man on the street's view" to be the only questionable premise of the argument, the conclusion is that the man on the street is in error . . . . But an argument whose conclusion is incompatible with its premises may be viewed in another way: as a heuristic, rather than a strictly deductive, argument. The major difference is that whereas the formal "conclusion" of a *reductio ad absurdum*

[6]cf. my above remark, that the timelike dimension of spacetime does not describe something through which anything actually moves, but is simply the denominator of the verb.

> is of no interest in itself (as being "absurd"), a heuristic argument can lead to the acceptance of a positive conclusion which, in some sense, *corrects* its premises. The principles of such argumentation are notoriously hard to codify; and it seems to me that, just for this reason, if one regards the process as a serious one—and not just a frivolous exercise or game—, one is obliged in pursuing it to seek the greatest possible clarity at all stages.

The reason for including this lengthy quotation here, is that it touches on a point in the epistemological argument that I've given in Chapter 2 (esp. § 2.4). For, the theories of Einstein and Parmenides really must both be incorrect for the same basic reason: they work from axioms that prove to be untenable because they lead to an absurdity. Therefore, it is the axiomatic basis of Einsteinian relativity that needs to be examined critically in order to reduce the theory to a new one which is not absurd.

Now, although Stein should therefore be entirely correct in his assessment of how the error comes into relativity theory, I believe he has made an error in pinning down what that error is. For there really should be nothing wrong with the 'man on the street's view', so long as he stands motionless with respect to all of his surroundings. The problem, I believe, occurs when we try to attach the same significance to the view of the 'man in the train', or the 'man in the cabin of the boat', who, although they are also inertial observers, cannot really be thought of as being in a true state of rest, any more than we think the Earth is, when we interpret the meaning of the dipole anisotropy in the cosmic microwave background (CMB). The fact that we now possess such *strong empirical evidence in favour of the existence of a true state of absolute rest*, as the CMB provides, suggests that such pre-cosmological arguments for the significance of the relativity of inertia, as Einstein employed in motivating his interpretation of the relativity of simultaneity, can no longer be given the weight that they once were. Instead, we should utilise—to its fullest extent—the fact that the empirical evidence suggests that there actually is an absolute frame of rest, as Hermann Weyl already realised in the early 1920s [20, 21, 22, 23, 24, 25, 26, 27], and aim for an interpretation of relativity theory that would describe that state consistently in local frames as well.

Accordingly, the Atomist-Newtonian-type interpretation of relativity theory that is described in Chapter 3 begins from the axiomatic standpoint there really is a four-dimensional void through which a three-dimensional material universe evolves with well-defined absolute cosmic time. I mean, that absolute cosmic time is a common property of every particle that makes up the universe, so that the present association of particles that may also move randomly with respect to each other, through the universe in cosmic time, evolves uniformly along the four-dimensional void.

This can be reconciled with special relativity theory according to the following kinematical principle: as the universe moves through the void, which in this case is four-dimensional Euclidean space, no particle in the equably evolving universe, which is always a well-defined hyperplane, can ever move further through the universe than it does in the direction of cosmic evolution. The map of occurrences in that well-defined three-dimensional universe is an evolving block spacetime that might somehow have been etched into the void, actually remaining in the past, as in the evolving block universe theory which has recently been discussed by George Ellis [28, 29, 30] (although there is no void in that framework); however, the past block of spacetime events really doesn't exist any more than the future one in the void/absolute cosmic time interpretation, as all that exists is the four-dimensional void and the present universe evolving in absolute cosmic time, carrying information about past occurrences along with it, e.g. in the form of photons which move at the cosmic speed limit.

The absolute cosmic time and speed limit are so far assumed axiomatically, as physical laws; and, in relation to the standard interpretation of special relativistic spacetime, the same properties of the physical description of spacetime events that did occur, can be subsequently recovered if (in the Euclidean space on which those events should be mapped) we simply replace the coordinate that describes the direction of flow of cosmic time with a pure imaginary number, so that what is described is an evolving block Minkowski spacetime, with its Lorentzian signature and everything that entails.

This interpretation of the physical description is similar to the one that was initially made by Hendrik Lorentz [31, 6] and Henri Poincaré [32]—which was, e.g., advanced later by John Bell [33, 34];[7] however, some nontrivial structure has been added to the description of physical reality, which further clarifies this position as one which is in fact consistent with the standard description of relativistic covariance, according to which the phenomena of time dilation and length contraction result purely from the finite speed of light through the universe and its invariant local measure. For we can now elaborate on the notion that the simultaneous events occurring as 'local time' progresses in an arbitrary inertial frame, do not necessarily *really* occur simultaneously, viz. in Lorentzian 'true time' [31, 6], by noting that at any instant in the frame of a relatively moving observer, the events that really do occur simultaneously (simultaneous noumena) take place on a spacelike hyperplane that extends to the 'past' and 'future' in the local coordination of four-dimensional *reality* at any instant, and, similarly, that the synchronous hyperplane extends into the void, where no universal noumena are occurring; i.e., there is an objectively well-defined distinction between the simultaneous phenomena that the theory describes as events in the four-dimensional spacetime continuum, which are coordinated differently by observers in relative motion, and those which are in fact simultaneous noumena. Then, as the special relativistic universe is clearly and objectively well-defined in all frames, it will be described, in the frame of an inertial observer who is not in a state of absolute rest, as evolving at some nonzero angle with respect to the plane of synchronicity, although the worldlines of all particles that travel through the universe at rates up to the cosmic speed limit still must be timelike, and the null worldlines describing the paths of photons will remain phenomenally invariant. Therefore, depending on the coordinate system that is used to describe the noumenal progression of events, the phenomenal measures of true lengths and durations will also transform covariantly.

Following Einstein, one might be tempted to label the absolute time and void of this theory as superfluous, since the description of phenomena is the same without them; but, lacking these additional elements, Einsteinian relativity theory is fundamentally at-odds with the common sense perceptions of Time and free will, and is therefore absurd. It should therefore be noted that this Lorentz-Poincaré interpretation of the theory is more objective, because, along with everything that is accounted for similarly as in Einstein's limited theory, it is essentially consistent with these most common facts of observation as well.

Another objection that might be made, is to the artificial introduction of Lorentzian signature into the special relativistic spacetime metric; however, this simply results from

[7] A recent volume, dedicated to the exposition of this alternate framework, is [35].

subsequently defining the local time axis as the worldline of an inertial observer who remains at rest in that frame, in conjunction with the requirement that the speed of light should be measured to have the same constant value in all such frames; i.e., by assuming the standard inertial and causal structures of the theory. The physics is thus already contained in the kinematical principles, while the metrical structure defined through the Wick rotation simply makes that easier to follow, i.e. through Minkowski's spacetime formalism. Therefore, that definition cannot be considered any more or less artificial than the Lorentzian signature that is generally assumed in relativity theory.

But because the Lorentzian signature of the metric may in fact be fundamental, as it would be if general covariance and the constant finite speed of light did really depend on it, the more important point to consider, is whether there is a more realistic way of formulating the theory, so that the Lorentzian signature should be described, instead, as a basic property of physical reality.

Together with this problem, it was then also important to consider whether the void itself should truly be dynamically warped by gravitational mass, just as spacetime is warped in frames describing the existences of massive bodies. The idea that I had to begin with, according to the cosmological SdS solution, was that the void should indeed be warped by a massive universe evloving along r from an initial singularity, with the curvature of the external (future) void given according to the Jebsen-Birkhoff theorem [36, 37, 38], so that the timelike nature of r would be forever induced according to the field theoretic description of the massive universe's interaction with the de Sitter void in which it was embedded.[8]

However, the eventual theory that I came to, although effectively equivalent to this idea, is essentially different.

The problem that I encountered right away, was that a timelike direction, in a metric with Lorentzian signature, is probably required to begin with, in order to describe the warping of spacetime around an existing mass. The reason (according to a well-known calculation) is simple: if one tries to describe the instantaneous curvature of the four-dimensional void around an ultimate point mass, assuming that it must be isotropic, the solution of Einstein's equation is simply the metric for maximally symmetric space.

This can scarcely be called a proof that gravitational mass shouldn't somehow actively warp the basic fabric of physical reality; and no doubt such a theory could be formalised some other way—i.e., some way that would not also require the block existence of the four-dimensional field in conjunction with that space, as in Ellis' evolving block universe theory, in which the past-and-present field that exists *is* the basic fabric of reality, which is warped around massive bodies as general relativity theory is supposed to describe;[9] but, under the

[8]The non-locality of this description seemed to be no less troubling than that of the standard black hole picture, in which the curvature of spacetime throughout $r < 2m$ should be induced by a future spacelike singularity that was already created in the principal collapse, to which all particles that subsequently fall into the hole are teleologically attracted.

[9]However, it should be noted that the only reason this would be possible, is that Ellis has forsaken the Einsteinian interpretation of the relativity of simultaneity as well, and assumed an absolute time so that the synchronous present in all frames is not also the noumenal present—for otherwise Einstein's argument in [5] proves a *reductio ad absurdum* in the same spirit as Putnam's: viz., that the standard idea that the spacetime field should really dynamically mould its accompanying space, taken in conjunction with Einstein's interpretation of the diffeomorphism invariance of that field, ends up requiring that the assumed dynamical

assumption that in reality there *is* a four-dimensional space with an independent ontological existence, there is also no good reason to *require*, just because our theory describes the progressively occurring spacetime field as being dynamically warped around gravitational mass, that this should be the case for the background void as well—especially when our theory also tells us other things, like the facts that *all* particles that exist must have energy (canonical conjugacy of E and t), that mass is equivalent to energy, that there is no effective difference between gravity and inertia, and that energy is the timelike component of the four-momentum.

When I realised that such points as these—combined with the fact that the supposed dynamical warping of the void around a gravitational mass is not naturally described through the reduced principles of the theory I was now attempting to synthesise—seemed to indicate that mass might actually be described as corresponding to the inertial path of a particle through the void, the answer to the question of whether there should be a basic Lorentzian signature in physical reality came very quickly, according to the strong empirical evidence suggesting a basic Cartesian ordering in Nature—since, according to this non-trivial fact, the void, now understood to be an absolute space according to Newton's definition,[10] should really have to be maximally symmetric.

Using Cartesian coordinates, these few spaces are straightforward to classify (see Eqs. (3.32) – (3.37), and the surrounding discussion): in n dimensions, they are either n-dimensional flat space, or n-dimensional surfaces with nonzero constant sectional curvature, which may be described as hyperspheres with constant radius of curvature, embedded in $(n+1)$-dimensional flat space; so, if we begin by assuming that the coordinates describing the maximally symmetric space should all be real, we find, e.g., that the flat, maximally symmetric real space must be Euclidean space, which correspondence to special relativity theory ammounts to a re-definition of one coordinate as purely imaginary, as discussed above. If the $(n + 1)$-dimensional flat embedding space is Euclidean, the only maximally symmetric hypersurface is the closed n-sphere. If, on the other hand, the embedding space is Minkowski space (by virtue of the embedded sphere that is described; see Eq. (3.34)), then the maximally symmetric space can have negative curvature, as described by the open n-sphere with imaginary radius of curvature (hyperboloid of two sheets), or positive curvature, as described by the open n-sphere with positive radius of curvature (hyperboloid of one sheet). The latter is de Sitter space, which is a real manifold with one timelike tangent vector at every point, so that its metric has Lorentzian signature; and the former, which is anti-de Sitter space, has a purely spacelike basis, so its metric is positive-definite.

---

warping of space doesn't actually take place in a fundamentally evolving way, but is moulded into a four-dimensional static block.

[10] sc. [3], 'Absolute space, in its own nature, without relation to anything external, remains always similar and immovable. Relative space is some movable dimension or measure of the absolute spaces; which our senses determine by its position to bodies; and which is commonly taken for immovable space; such is the dimension of a subterraneous, an aerial, or celestial space, determined by its position in respect of the earth. Absolute and relative space are the same in figure and magnitude; but they do not remain always numerically the same. For if the earth, for instance, moves, a space of our air, which relatively and in respect of the earth remains always the same, will at one time be one part of the absolute space into which the air passes; at another time it will be another part of the same, and so, absolutely understood, it will be continually changed.'

The proof of this coherent description, which is given in § 3.3, is significant for a few reasons. For one thing, it is commonly said that the open hyperbolic slice of Minkowski space with negative curvature, is the negative curvature analogue of the closed sphere—i.e., that the geometries described by

$$ds^2 = \frac{dr^2}{1 - Kr^2} + r^2 d\Omega_{n-1}{}^2 \tag{4}$$

are (closed) spherical or (open) hyperbolic, depending on whether the curvature constant K is positive or negative, respectively. For basic topological reasons, such a correlation actually should not be drawn.

For example, when $K > 0$, there really is no good reason for assuming that the coordinate r goes only from 0 to K, and that the space that the line-element *really* describes is the closed sphere which continues on in a regular manner below the equator, as described, e.g., in [39]; for it is just as true that for $1/\sqrt{K} \leq r \leq \infty$, the line-element describes de Sitter space—just as the curve $y = \sqrt{1 - x^2}$ is hyperbolic, with y purely imaginary, for $x^2 \in \mathbb{R} \geq 1$; i.e., it is a one-dimensional surface with real domain, embedded in the complex plane.

And, furthermore, if we should be daunted by the complexity of this embedding space, then we must be so consistently, and say that we don't like the curve $y = \sqrt{-1 - x^2}$, $x \in \mathbb{R}$ either. But we don't do this:—we describe this very curve—or its higher-dimensional counterpart, given by rotations about the (imaginary) y-axis—as open hyperbolic space with negative curvature; and we call it the analogue of closed spherical space with positive curvature; and we describe de Sitter space as a maximally symmetric pseudo-Riemannian manifold, with Lorentzian signature and positive curvature.

However, the maximally symmetric space with negative curvature, or disconnected open sphere (hyperboloid of two sheets) with imaginary radius of curvature embedded in (induced) Minkowski space of one higher dimension, is not the analogue of the closed sphere with positive radius of curvature, but of de Sitter space, which may be described as a *real* maximally symmetric space with positive curvature, or a connected open sphere (hyperboloid of one sheet) with positive radius of curvature embedded in Minkowski space of one higher dimension. And the latter is every bit as *Riemannian* as the former.

And, for what it's worth, it may also be noted that the closed sphere should not necessarily be described as a surface embedded in Euclidean space, but may also be embedded in de Sitter or anti-de Sitter space as well; i.e., spherical space may be described as a maximally symmetric closed hypersurface of any one of these three open space forms.

And according to this coherent description, since the basic metric tensor of anti-de Sitter space is positive-definite, it should be noted that any description of an anti-de Sitter spacetime requires one coordinate of the basic metric to be defined as imaginary, in the same *ad hoc* manner as is done in constructing special relativistic spacetime.

Now, this is really all very significant because of the fact that Einstein continually claimed that general relativity theory would be logically simplest if $\Lambda \equiv 0$. However, if we consider that the Lorentzian signature of spacetime in relativity theory should be related to a basic property of Nature, so that Physical Reality would *essentially* give rise to a generally covariant description, then it is really logically the simplest—whether or not we should admit the existence of a void into our physical theory, or whether such a void, if admitted, should be deformed by any absolute rest-mass that moves through it, or should itself be an absolute space; i.e., in any case—to assume a positive cosmological constant in Einstein's equation, so that its most basic solution should describe a real four-dimensional Riemannian manifold with intrinsic Lorentzian signature.

In fact, it is important here to briefly describe, as best I can in a few short words, one (arguably the only) negative aspect of Einstein's epistemological method. In order to obtain the greatest logical simplicity, he felt that one should begin with a physical theory, and sort out how it may have been derived according to the fewest number of axioms (or mathematical terms); and then cast aside, as superfluous, all others that might initially have been thought to be realistically assumed, before proceeding to refine the theory in order to bring it in line with even more facts of observation, taking care to admit only those new principles or axioms which seem to be the most reasonable; then, with such a reasonable and logically simple theory in hand, he should work to develop an acceptable level of comfort when certain aspects of the theory don't naturally make sense, as paradoxical inferences may be encountered.

But there is a fault in this, as he did search principally for logical simplicity in the axioms of his physical theory, even at the sake of some loss of intuitive clarity at the fundamental level: for while the historical development of physical theory does strongly suggest that the correct theory should be logically the simplest, no one is fit to judge, while speculating about what the principal axioms of that theory should be, which ones will give rise to the logically simplest theory that may subsequently be deduced.

For example, while he did not immediately realise that his interpretation of the various coordinate representations in special relativity theory should amount to the description of a block universe, he did eventually realise and accept this fact, which, although there is no definitive means of proving that it should not truly be, still seems extremely unrealistic. Another paradoxical aspect of Einstein's theory is noted in his autobiography, where he wrote that in developing general relativity theory, after he had realised the principle of general covariance in 1908, he could not complete the thoery until he had spent seven difficult years convincing himself that the coordinates should have no immediate metrical meaning [40]. The same level of comfort with this paradoxical notion is carefully developed, e.g., at the beginning of Charles Misner, Kip Thorne, and John Wheeler's *Gravitation* [41].

In contrast, the coordinates do have immediate metrical meaning in the theory developed in this thesis. And this begins by framing the de Sitter manifold with a particular Cartesian coordinate basis, which sets it up as the absolute background space through which the three-dimensional universe should really evolve, and the ideal evolving block spacetime should emerge. Then there is only one realistic option for the description of a three-dimensional universe that evolves in a causally coherent manner: the comoving, torsionless closed 3-sphere which evolves in the timelike direction of the corresponding five-dimensional Minkowski embedding space. This is contrasted with more usual coordinations of the de Sitter sphere—i.e., the statical Eddington frame [1] and the comoving Lemaître-Robertson frame [42, 43]—towards the end of § 3.3, in an analysis which clearly illustrates the illegitimacy of those descriptions as realistic cosmological models, so long as metrical significance of the coordinates is granted and de Sitter space is not conceived as an abstract manifold.

Now, an interesting aspect of this universe, as opposed to the universe of special relativity theory, is that the only inertial particles—i.e., which move with constant velocity through it, as cosmic time progresses—whose paths are actually geodesics of de Sitter space, are those which remain perfectly at rest. All other inertial particles are accelerated through the universe. This was already realised by Henri Bacry and Jean-Marc Lévy-Leblond in 1968 [44], who named this particular Lie group the *Newton group*, but failed to realise that the accelerations of these inertial particles might, in their proper frames, be perceived as mass.

It is worthwhile to briefly discuss the content of Bacry and Lévy-Leblond's paper, because, along with the fact that it corroborates the above argument, that $\Lambda > 0$ is really the logically simplest possibility for general relativity theory, the formalism that they develop will be useful for outlining the rest of the thesis in a clear and concise way.

First and foremost, what Bacry and Lévy-Leblond proved [44], was that there are eight different types of Lie algebras for kinematical groups that one can reasonably propose, and that the *de Sitter group* corresponds to the most general of these, so that all the others can be recovered as limiting cases of it, according to an Inönü-Wigner 'contraction scheme' [45]. The article was highlighted a few years later by Freeman Dyson [46], who noted, e.g., that the de Sitter group is the only simple group among these.

The first of the two types of contraction which are relevant here, is the so-called 'speed-space' contraction, whereby the generators of space-translations and inertial transformations are considered in their infinitesimal limits. This contracts the relative time Lie algebras to the absolute time Lie algebras; e.g., the speed-space contracted de Sitter group is the Newton group, which is relevant for us, and will be discussed further shortly.

The other interesting type of contraction is the 'speed-time' contraction, in which the time-translation generator and inertial transformation generators are sent to zero. Just as the relative time Lie algebras are contracted to the absolute time Lie algebras through speed-space contractions, so the relative space Lie algebras are contracted to absolute space Lie algebras through speed-time contractions—which describe spatial slices, in some particular proper inertial frame, as being similar for all times. Bacry and Lévy-Leblond wrote that this process 'only yields groups describing intervals connecting events without any causal connection, hence without much physical applications' [44].

However, I think this was too hasty a conclusion, since the contraction doesn't necessarily need to be taken to mean that *nothing* can move through space (i.e., that we have to set $c \to 0$ and thereby refuse to describe the motions of test-particles), but merely that space remains absolute relative to *something* that is always described as being at rest;[11] in fact, this doesn't even need to mean that that something is *actually* at rest, in an absolute sense. However, it is relevant to note that such a description of space cannot pertain to a situation in which its geometry is generated by a local action causally emanating from a massive body.

In relativity theory, one very important group to which this contraction pertains, was already noted in a way by Eddington in 1923 [1], when he sought to describe the relatively symmetrical properties of an ultimate particle of matter, about which photons could be described to move, in the 'radial' direction, 'forever' at the same speed—i.e., so that the surrounding spacetime description could be locally restricted in algebraic form, to

$$ds^2 = -U(r)dr^2 - V(r)(r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2) + W(r)dt^2. \tag{5}$$

[11] Analogously, in the case of speed-space contractions, we don't actually have to say that the absolute time that they define is also the proper time of *all* observers, and thus the same time-coordinate of all frames. This is why we can still reasonably define an absolute time in special relativity theory, contracting Einstein's *Poincaré group* representation to the Lorentz-Poincaré *Galilei group* representation of the theory.

This is Eddington's equation (38.12), as he referenced it in § 66 of [1]. In fact, because this corresponds to a speed-time contraction, U, V, and W should most generally be written as functions of t as well, since the line-element should describe the non-causal curvature of space surrounding a particle at rest at some time, and *similarly* for all times; however, that potential t-variability can be gauged away when the metric is required to be a solution of Einstein's equation, according to the Jebsen-Birkhoff theorem.

In Chapter 4, the local SdS solutions to Einstein's equation are analysed according to the interpretation that the line-element corresponds to such a speed-time contraction of the de Sitter spacetime field surrounding an ultimate particle of matter moving with arbitrary inertial velocity through a framed de Sitter manifold. More will be said about that analysis shortly, which should be the most clear after the relevant speed-space contraction has been discussed.

By sending the space-translation and inertial transformation generators to zero, speed-space contractions distinguish a particular bundle of worldlines of particles that coherently evolve *as* a (three-dimensional) universe with absolute time, which therefore satisfies Weyl's postulate. Such contractions of the de Sitter group are the easiest to conceptualise if it is thought of as a framed manifold:—the four-dimensional hyperboloid of one sheet, embedded in five-dimensional Minkowski space. The trivial contraction, which is analogous to the pure de Sitter form of Eq. (5), distinguishes the inertial *geodesics*, which actually do remain at rest in the framed manifold, as the fundamental worldlines. The corresponding Robertson-Walker (RW) metric, describing the evolution of this universe—i.e., describing this Newton group—from this perspective, is

$$ds^2 = -dt^2 + \frac{3}{\Lambda}\cosh^2\left(\sqrt{\frac{\Lambda}{3}}t\right) d\Omega_3{}^2, \tag{6}$$

where $d\Omega_3$ describes the cosmic 3-sphere which contracts until $t = 0$, and expands thereafter.

However, since it has already been noted that these worldlines describe the only non-accelerated inertial paths in the universe, it seems most reasonable to suppose that a particle following such a path should be massless; indeed, according to the statical Eddington line-element, given by the speed-time contraction around such a geodesic, a particle that exists as such, remaining forever at $r = 0$ in

$$ds^2 = -\left(1 - \frac{\Lambda}{3}r^2\right)c^2dt^2 + \frac{dr^2}{1 - \frac{\Lambda}{3}r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{7}$$

should be massless.

Therefore, the proposed cosmological framework is given as follows: inertial geodesics of the framed de Sitter sphere are paths of massless photons; the fundamental rest-frame of massive particles is the collection of worldlines that actually describe particles which move at the rate of null-lines relative to these photons—i.e., the null-worldlines themselves; in particular, we ustilise the Hopf fibration of the 3-sphere ($SU(2)$) to identify two opposing tangent directions at each point on a 2-sphere ($SU(2)/U(1)$; with tangent vectors directed either way along the circle ($U(1)$) at each point), and define the fundamental massive particles as those which move all in the same direction at the null speed; accordingly, we write down the metric for the 3-sphere which expands in cosmic time, so that photons are now described as travelling along null-lines in that direction, as

$$ds^2 = -B(r,t)dr^2 + A(r,t)dt^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{8}$$

where r is the timelike coordinate describing the radial expansion of the actual universe in this frame; but then, by definition, the null-lines of this system must be both, the fundamentally inertial geodesics *and* the inertial worldlines of particles that move through the 3-sphere at twice the fundamental null-rate in that direction, so that particles travelling along these perceived null-lines should also be identified as photons *by association*; and finally, the picture is completed by assuming continuous symmetry, by which the homogeneous and isotropic universe would be filled uniformly with particles moving with all possible velocities, so that throughout the 3-sphere there will be as many particles with velocities pointing in one direction, as there are particles with velocities pointing in the other, and those velocities may in fact range in magnitude from $-\infty$ to $+\infty$.

According to a massive particle travelling with velocity between the two null-lines defined in Eq (8), the paths of all other particles with similar velocities will be described as timelike worldlines. Also, photons travelling along fundamental geodesics will be incoming, while those travelling at twice the rate of a cosmic rest-frame observer will be outgoing; therefore, the coordinate t in Eq. (8) is associated with the proper radial direction in massive frames. This description is equivalent for particles with velocity in the opposite direction of the circle as well, which associate with the bundle of null-lines propagating in the opposite direction, thereby defining a third set of worldlines pertaining to photons (i.e., particles with twice the null-velocity in that direction). These two types of matter should therefore be quite similar, but because their relative velocities are greater than those of photons, they must also be relatively very different.

This difference is realised as follows: all of these velocities correspond to radial momenta, which enter into the statical line-elements, given by the speed-time contractions (the local form of Eq. (1), which is found by restricting Eq. (5) to a solution of Einstein's equation), as mass-terms; the mass of a (spherically symmetric, uncharged) particle, which is identified as precisely half the horizon radius, is not a single value, but a triplet—e.g., a photon should not merely have trivial mass, which is the mass of an incoming photon in every frame, but $2m_\gamma = \{0, \pm\sqrt{3/\Lambda}\}$, corresponding to the three paths defined as null-lines through the two distinct massive frames; accordingly, each massive particle has either two positive and one negative mass, or two negative and one positive mass; the former are described as positive mass particles, and the latter as negative mass particles, and these are mutually invisible and gravitationally repulsive, as determined in § 4.2.

It is significant, that each triplet, or superposition of three mass values, is as close to an objective definition of mass as can be given by the geometry: a particular value fixes the other two unambiguously; and the geometry, from which the mass is determined, pertains to this degenerate set. This can be understood kinematically as follows: due to the symmetrical properties of the description resulting from the conjunctive definitions, of actual null-lines as the fundamental rest-frame geodesics, and the actual geodesic paths of massless particles with trivial inertia as null-lines, the inertial speed of a massive particle through the universe, say $0 < (d\vartheta/dt)_m < (d\vartheta/dt)_{\rm null}$ in the coordinate system of Eq. (6), must be relatively equivalent to that of a particle with actual speed $2(d\vartheta/dt)_{\rm null} - (d\vartheta/dt)_m$. But also, from the perspective of this same cosmic rest-frame, the relative speed of a particle with 'negative' mass, $-(d\vartheta/dt)_{\text{null}} < -(d\vartheta/dt)_{-m'} < 0$, should actually be equivalent to that of a particle *moving in the opposite direction*—i.e., the same direction as 'positive' mass particles—with speed $2(d\vartheta/dt)_{\text{null}} + (d\vartheta/dt)_{-m'}$ greater than the null-rate prescribed for photons in that cosmic rest-frame. Then, by extension, relative speeds between $-(d\vartheta/dt)_{\text{null}}$ and $-2(d\vartheta/dt)_{\text{null}}$ should correspond to particles with negative mass, and those less than $-2(d\vartheta/dt)_{\text{null}}$ should correspond to particles with positive mass. Accordingly, the mass of a particle is a degenerate triplet corresponding to its fundamental radial inertia (and the local SdS solutions, by utilising the relativity of inertia in the appropriate speed-time contraction, describe spacetime around particles that remain at the origin, so that this inward or outward inertia corresponds to rest-mass), and further determination than that is not possible, relatively speaking.

Therefore, the inertial velocity of a particle through the universal 3-sphere, which was used in order to determine the cosmic rest-frame for massive particles, should not be thought to correspond to a cosmological picture in which some particles 'move this way', while others 'move that way' through apparent space, but simply to each particle's rest-mass, as it exists in the fundamental space with non-vanishing 'radial inertia', which is absorbed into the timelike coordinate in its proper speed-time contraction.

This is really the most intuitive way of describing cosmological test-particle dynamics of the fundamental Newton group, Eq. (6); as each massive observer, with non-trivial inertial velocity through the fundamental 3-sphere (although they should all naturally consider themselves as being at rest), should naturally coordinate its metric, at any time, as

$$ds^2 = d\vartheta' + \sin^2 \vartheta' d\Omega^2, \tag{9}$$

so that ϑ' is the radial coordinate, which is locally like the radial coordinate of flat space, and space is conceived by such an observer, with non-trivial velocity in the (positive *or* negative) ϑ-direction, as infinitely many 2-spheres stacked along this dimension, containing matter which might be relatively moving through those three dimensions as well, which should each personally conceive of their own surroundings in the same way, so that each contracts their local relativistic spacetime coordinate system according to this idea of radial symmetry in proper inertial space, along with an assumption on the proper radial velocity of photons, so that the fundamental velocity of each particle must enter into the metric in another way.

With this description in mind, we return to the cosmological picture. First of all, note that the Newton group, which is trivially described by Eq. (6), is a 3-sphere which contracts to a finite size at $t = 0$, and expands afterwards. However, by defining a 'radial' timelike-coordinate in the manner specified in Eq. (8), in order to contract this same group about a bundle of null-lines of de Sitter space, the metric was constructed as one which should only describe the expanding *or* contracting half of the group. Thus, the expanding cosmological description in the coordinates given by the eventual solution, Eq. (1), is singular at $r = 0$, where expansion begins, whereas the contracting half collapses to this singularity. Interestingly enough, the mass in the cosmological solution actually is a single, well-defined value.

Now, in the proposed large-scale scenario, this expanding universe is supposed to begin with a homogeneous distribution of positive and negative mass particles, which naturally dissociate, through mutual gravitational repulsion, into a sort of lattice distribution of clusters of matter with positive and negative mass, which are mutually invisible. This repulsion will not, however, contribute to the cosmic rate of expansion, which is a well-defined geometric property. Instead, the gravitational repulsion that would be dynamically felt within any massive cluster, due to the presence of an effective local shell of 'negative' mass, would be equivalent to an additional centrally attractive force within the cluster.

This theory for the true nature of locally non-interacting dark matter and natural formation of the void-filament structure observed in our Universe, as well as observed baryon asymmetry, is quite similar to a recent proposal by Massimo Villata [47, 48], except that he has suggested that the repulsive gravitational interaction between matter and anti-matter should contribute to the cosmic expansion rate. In fact, the theory is even consistent with Villata's idea that these anti-matter particles should, in one way, be describable as moving backwards in time, even though we now also describe everything as remaining within the three-dimensional universe, moving forward in its absolute cosmic time.

And as this occurs, the global expansion rate of this RW universe should be determined from the cosmological SdS solution, Eq. (1). For the universe really is, by construction, homogeneous and isotropic; and although it is not synchronous in the proper coordinate frame of massive particles in the cosmic rest-frame, it must appear to them as such, because those non-trivial fundamental worldlines do fan out isotropically as the universe evolves, so that relative to any one, all others would appear to be isotropically receding. Furthermore, the apparent rate of cosmic expansion must be given according to the change in r, as measured in the proper time coordinate system of the fundamental rest-frame, Eq. (2), which is the same as the flat ΛCDM scale-factor.

The connection to FLRW theory is then trivial: in that theory we assume a universe which is isotropic and homogeneous, and expands in absolute time at a rate which is determined by three parameters—the density and pressure of the various forms of matter it contains, and its spatial curvature; the form of the scale-factor which describes this expansion is constrained, through Einstein's equation, after assuming perfect fluidity of matter as an acceptable large-scale approximation, to a solution of the two Friedman equations, which depend on these three physical parameters; we then solve this under-determined system of equations by considering equations of state, relating pressure and density for various types of perfect fluid that the universe might contain, and subsequently model observables in order to constrain the form of the scale-factor, which depends on the values of the three physical parameters; accordingly, observations have constrained the scale-factor quite well to the form of Eq. (2); the theoretical universe described in this thesis must bear that very same appearance, since it appears from the fundamental rest-frame to be expanding isotropically, and scaling, according to the proper measure of the expansion factor in that frame, according to Eq. (2).

Some further points to consider are: the expansion occurs for purely geometrical reasons, rather than due to the local dynamics of material in causally connected regions of the universe, so there is no horizon problem; the apparent scale-factor is precisely that of the flat ΛCDM universe for all time, so there is no flatness problem because any measurement of cosmic expansion must constrain the curvature parameter to zero; the prior resolution of both of these problems, together with the fact that the rate of evolution is well-defined for all time, means that there is no need for an inflationary epoch; in fact, because the cosmic expansion is not driven by the local energy content, but is a well-defined prior property of the universe and the underlying void, there is also no cosmological constant problem, because the vacuum expectation value from quantum theory simply cannot contribute to the expansion rate.

But beyond that, the geometry of this universe is essentially similar to the one that describes a locally spherically symmetric mass distribution. Furthermore, as shown in the thesis, the radial coordinate in the local geometry actually should not exist on the supposed 'timelike intervals', so that causally coherent, spherically symmetric gravitational collapse beyond the horizon radius, which cannot occur until the end of cosmic time, must not be described by that solution. Instead, by hypothesising that spherically symmetric gravitational collapse in one universe should, at the end of its cosmic time, produce a contracting Newton group, the picture is completed as one which is not merely consistent with cosmology, but should actually provide for itself the basic cause for the cosmic time that needs to be assumed, because this new universe might also be described by the contracting cosmological SdS solution, with particle dynamics described in the same way, which would eventually reach $r = 0$ in those coordinates, at the equator of the de Sitter sphere that was framed at the final stage of local collapse in the progenitor universe, and then emerge from that singularity, in a hot state following complete gravitational collapse, to be described now by the expanding cosmological SdS solution. Thus, a connection is made between big bang universes like ours and gravitational collapse, which should finally account for becoming in this theory.

And so, the thesis, which is sometimes deeply speculative, should not be accepted purely according to its philosophical appeal—viz., as it is very intuitive and common sensical, and admits only the most reasonable and globally consistent possibilities when confronted with dilemmas,—but because this leads to a metrically well-defined, fundamentally covariant theory of Physical Reality which does not just fix, but essentially resolves the most significant problems of the current dynamical theory, providing logically consistent, realistic answers for the cause of the Big Bang and the basic Nature of Time through causally coherent gravitational collapse, according to a theory which does not require any censorship or protection conjectures to resolve paradoxes of black hole creation in finite cosmic time or time travel to a past that does not exist, and leads naturally to a theory for the origin and evolution of the large-scale structure we observe in the Universe, in a continuously symmetric cosmology describing a universe that does indeed expand for all time purely due to Λ, with evolution which would be observed to have taken place by all fundamental observers precisely according to the flat ΛCDM model that has been empirically constrained.

The development of this theory is presented as follows. The first chapter is an examination of the historical development and current status of modern cosmology, beginning from the initial evidence for cosmic expansion and the theoretical modelling that became possible with the advent of general relativity theory; i.e., the era of modern cosmology is defined as having begun in the years 1912 – 1917. Special attention is paid to the theories surrounding Λ, with an eye kept to the development of a new theory which would be consistent with the principles used to establish RW spacetime as a cosmological background geometry, but also consistent with Eddington's interpretation, that cosmic expansion might be objectively required by a geometrical property of Physical Reality. Therefore, after outlining the details in the history of early modern cosmology, from which the RW line-element emerged as providing the general description of an isotropic and homogeneous expanding universe (§ 1.1), the reasoning that was given for the two competing interpretations of expansion is discussed in § 1.2, in order to establish the case for Eddington's interpretation as a more realistic alternative to Einstein's.

Then, various ways in which the cosmological constant was subsequently incorporated are discussed in § 1.3, as these eventually led to the interpretation of a dark energy component of the Universe after its existence was finally discovered in 1998 [49, 50].

Now, the big bang FLRW models are not only inconsistent with the interpretation of Λ as the basic reason for cosmic expansion, as they do lend themselves more naturally to the description of Λ as a dark energy component of a universe that resulted from a singularity (unless the theory of gravity itself is modified in order to avoid this), but there are other significant problems, such as the observed flatness, isotropy on the largest scales—which, together with the cosmological principle, implies large-scale homogeneity (isotropy from every point), but also a horizon problem since this could not result from any causal influence in the universe that is described just so,—and the fact that all the matter we know anything about, which we can directly observe, makes up only four percent of the measured total energy in the Universe. Actually, the most significant of the cosmological problems is likely the paradoxical implication from relativity theory, that the Universe may not be evolving at all, but that each present instant might be instead just one subjective slice of a four-dimensional Block Universe.

But the recurring theme here, especially in the more well-known problems in cosmology (horizon, flatness, dark matter/dark energy), is not that there is any discrepancy between the observational data and the physical model, but that the empirically constrained parameters are inconsistent with our expectations of what should realistically be. Therefore, the remainder of the thesis is an attempt to revisit those expectations, and derive a theory from rationalised first principles which would be consistent with the evidence: that the Universe is expanding, and probably due to a cosmological constant, with a measurable expansion rate equivalent to the description in which the (other) dominant energy component indicates the presence of an appreciable amount of invisible dust; that it began expanding from a hot initial state; and that it must appear perfectly isotropic and flat to all observers. Thus, the theory would be observationally equivalent to the standard (flat ΛCDM) concordance model, but would provide reasons why things ought to be the way they appear.

This begins, in Chapters 2 and 3, with an examination of the physical nature of time, which is essential to the problem of deriving a nontrivial relativistic cosmological model. Therefore, these chapters mainly present an argument for why relativity theory in general can and must be interpreted as describing the evolution of a three-dimensional universe, in a manner that is consistent with, but generalises the trivial prescription of an absolute cosmic time that is made in deriving the RW line-element; thus, e.g., it is meaningful to say in general, that a particle which has existed in the Universe since the Big Bang has *really* 'been around' for the past 13.7 billion years, even though it may have been travelling all that time through the cosmic rest-frame at some significant fraction of the speed of light so that its relative proper time would be considerably less; i.e., the purpose of this argument is to generally reconcile relativity theory with the sense of the real present that relativistic cosmology takes for granted.

So, beginning with an investigation into the problem of the physical nature of time in its historical context, the argument proceeds from an epistemic standpoint. The purpose of Chapter 2 is therefore to discuss the foundational insights that were made in the philosophy of time, from Heraclitus and Parmenides, through the Atomists, Aristotle, and St. Augustine, eventually to Newton and Einstein, and to discriminate between the epistemological methods that developed along the way, in order to form a case against the unnatural inferences that were finally made by Einstein and Minkowski. As such, when discussing the theories of Heraclitus (§ 2.2) and Parmenides (§ 2.3), along with the reasoning behind them, a great deal of consideration is given to the epistemological methods that they used; and this eventually leads, in § 2.4, through further discussion of how those methods were incorporated into those of their successors, to the formulation of the natural epistemic stance that is taken in this thesis. In § 2.5, the Atomists' theory is presented as one based on a reduction of Parmenides' argument, and then Augustine's recognition that time must be equable duration (as Newton eventually put it), is discussed in § 2.6, before finally coming to the interpretations of special relativity theory by Einstein and Minkowski, in § 2.7, and their abandonment of the significant philosophical inferences that had previously been made with regards to the Nature of Time.

The chapter therefore sets up the argument, that Einstein, by simply rejecting Newton's absolute time, space, and motion, and taking the positivist stance regarding the relativity of simultaneity, but the stance of a realist when interpreting the consequential meaning of that inference, was compelled to agree with Minkowski that we must live in an absolute world like that of Parmenides, when he should rather have reduced Newton's philosophically well-motivated theory to one that would be consistent with the causal and intertial structures of the spacetime metric in special relativity theory.

Chapter 3 then proceeds with an attempt to reconstruct relativity theory as a truly dynamical one, consistent with the principles of Heraclitus, the Atomists, Augustine, Galileo, and Einstein, with the idea that there should be a four-dimensional void, acting as a general relativistic *Führungsfeld* through which the three-dimensional universe should evolve, with events in spacetime emerging with the necessary metrical structure. Therefore, the past and future events of spacetime are described as being *ideal* occurrences in a truly evolving three-dimensional present. The case for special relativity is neatly illustrated in § 3.1, with the example of a barograph in which a number of pens exist in a pre-arranged line with their paths traced out on graph paper that moves underneath by clockwork; and a reassessment of the objectively real potentialities that might give rise to the same empirical model (viz. special relativistic spacetime) is presented in § 3.2, with the eventual conclusion that this alternative scenario should have the best chance of leading to a determination of the essential nature of time. At the end of this section, an argument against the causally coherent formation, and present existence of black holes in our Universe begins to develop. Then, in § 3.3, the curvature and geometrodynamics of spacetime are considered, and it is eventually hypothesised that these might actually be described as relatively projective properties, purely due to particle dynamics in de Sitter space, rather than as the result of a dynamically warping *Führungsfeld*.

The new cosmological theory really begins to build from this point; and, in the latter half of § 3.3, upon the assumption that the absolute cosmic time should be given according to the induced metric of a de Sitter sphere, as an absolute background structure with intrinsic Lorentzian signature, two line-elements—one local, and one cosmological—emerge from considerations of how the events of four-dimensional spacetime might be algebraically coordinated. These are the SdS metrics—and the two physically different, yet algebraically equivalent statical geometries are described in § 4.1, using dimensionless coordinates normalised by the cosmical quadric of curvature, according to their fundamental parametrisation. In § 4.2, the description of physical mass in the geometry is considered, and a possible implication for the nature of dark matter and the formation of large-scale structure in the Universe is discussed. Then, in § 4.3, the cosmographical metric is analysed, and it is shown that the factor of expansion in the essentially isotropic and homogeneous universe would be perceived, by any observer in the fundamental rest-frame defined by a bundle of worldlines satisfying Weyl's principle, to evolve in proper time precisely according to the flat ΛCDM scale factor.

It is significant to note—with reference to my basic hypothesis, that a different cosmological ontology should be able to be given so that we might rationally *expect* the observation of flat ΛCDM-type expansion,—that in this scenario the bundle of fundamental worldlines are not orthogonal to the isotropic and homogeneous cosmic hypersurfaces, which is one of the basic assumptions that Robertson did make, and is therefore one of the basic mathematical principles of FLRW cosmology [27]. In fact, the fundamental worldlines in this theory are not geodesics of the fundamental background structure that is assumed—and they are not even timelike, but null;—rather, the bundle of timelike geodesic worldlines that are orthogonal to the cosmic hypersurfaces (in which frame the comoving spacetime line-element *is* RW) should actually be massless, e.g. with $M = 0$ in their corresponding local statical (S)dS geometries.

Finally, because this cosmological model, which begins from a coordinate singularity of the relevant metric, is described only on the expanding half of the de Sitter sphere, it is pointed out, in § 4.4, that because the corresponding description on the contracting half of the sphere is relevant to the description of the final stage of spherically symmetric gravitational collapse, it may be possible to describe the emergence of such a universe as the result of gravitational collapse in a prior one, which process would give rise to the absolute time that has to be assumed in the model.

Chapter 1

Modern Cosmology and the Cosmological Constant

> It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in the nascent state.
>
> —James Clerk Maxwell, *A Treatise on Electricity and Magnetism*

1.1 Early Modern Cosmology

The era of modern cosmology, in our continued search for an accurate understanding of the large-scale structure and evolution of the Universe, began only a century ago, when advances were made in physical theory due to Albert Einstein's general relativity theory [51, 6], and, shortly thereafter, to the observations of cosmic expansion. This expansion was first evidenced by redshift measurements of spiral nebulae, after the task of measuring their radial velocities was initiated in 1912 by Vesto Slipher [52]. Within a decade, the compiled list of these measurements, being predominantly recessional, strongly evidenced the expansion of the Universe [1]. The expansion was eventually confirmed by Edwin Hubble, who, after using observations of Cepheid variable stars to show that the spiral nebulae were distant galaxies, in 1926 [53, 54], determined, in 1929, a linear velocity-distance relation among the nearby galaxies [55].

Prior to Hubble's discovery, few models had been proposed as potential descriptors of the Universal geometry. In particular, the static, homogeneous, finite geometry proposed in 1917 by the father of general relativity theory [56, 6], required a modification of the original Einstein equation,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \kappa T_{\mu\nu}, \tag{1.1}$$

by the addition of a cosmological term,

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}, \tag{1.2}$$

in order to balance otherwise contracting physical distances between fundamental observers in the matter-filled Universe; i.e., 'necessary ... for the purpose of making possible a quasi-static distribution of matter, as required by the fact of the small velocity of the stars.' The constant Λ in Eq. (1.2) is called the *cosmological constant*, and the Λ-term is the only term up to second order which may be added to Einstein's original theory, Eq. (1.1), without violating general covariance, and thus local energy conservation.

An alternative to Einstein's static solution for the Universe was found by Willem de Sitter in the same year that Einstein published his model, in a successful attempt to discover a solution with a 'preferred' elliptical space, as opposed to the spherical space in Einstein's solution [57]; however, as Arthur Eddington noted a few years later, this discrepancy between the two spatial geometries is really due to an arbitrarily introduced division of space and time, and that Einstein's *spacetime* geometry is cylindrical, with time uncurved, while de Sitter's is formally equivalent to a 4-sphere, say of radius R, as its line-element may be written,

$$ds^2 = R^2[d\omega^2 + \sin^2\omega\{d\zeta^2 + \sin^2\zeta(d\theta^2 + \sin^2\theta d\phi^2)\}], \tag{1.3}$$

where ω, ζ, θ, and ϕ are four angular coordinates [1].

De Sitter's solution contains no matter, and had therefore been a blow to Einstein who had hoped Mach's principle, along with the principle of relativity expressed by general covariance and the principle of equivalence, would be one of three cornerstones of the correct theory of gravitation; i.e. he hoped the geometry would be completely determined by the stress-energy of matter [58]. Abraham Pais [58] writes,

> Einstein never said so explicitly, but it seems reasonable to assume that he had in mind that the correct equations should have no solutions at all in the absence of matter. However, right after his paper appeared, de Sitter did find a solution ... with $\rho = 0$. ... In 1918 he looked for ways to rule out the de Sitter solution, but soon realised there is nothing wrong with it.

De Sitter's seminal paper of 1917 [57] contains the first attempt at a relativistic interpretation of the observed velocities of the spiral nebulae, previously evidenced by their redshifts. Here, he analysed the static solutions of Einstein's equations for which the matter was supposed a pressureless perfect fluid, denoted *dust* in modern parlance, with constant stress-energy $T_{00} = \rho$, else $T_{\mu\nu} = 0$, and found that if the density was negligible (in fact, he set $\rho = 0$) the equations, Eq. (1.2), are satisfied by

$$ds^2 = -\cos^2\chi c^2 dt^2 + R^2 d\chi^2 + R^2\sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2), \tag{1.4}$$

with $\Lambda = 3/R^2$. De Sitter compared his solution to the Minkowski and Einstein geometries, which also satisfy Eq. (1.2) with the given stress-energy, and found that if his geometry would be the correct representation of World-structure we should observe redshifts in the spectra of distant objects as 'the frequency of light-vibrations diminishes with increasing distance from the origin of co-ordinates' due to $g_{00} = \cos^2\chi$, so that the time for a cycle of vibrations would be proportional to $\sec\chi$. He therefore proposed that this redshift might be measured in the spiral nebulae, which were not yet known as distant galaxies similar to the Milky Way, but, due to their measured radial velocities, were nevertheless thought to be the most distant objects known. In 1917, de Sitter was able to quote only three reliable radial velocity measurements of the nebulae, which gave 2 : 1 odds in favour of his prediction.

Later, Eddington further analysed de Sitter's solution and found a change of basis to the static, spherically symmetric line-element [1],

$$ds^2 = -\left(1 - \frac{\Lambda}{3}r^2\right)c^2dt^2 + \frac{dr^2}{1 - \frac{\Lambda}{3}r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{1.5}$$

This solution had in fact been found in 1918 by Friedrich Kottler [59], who generalised Karl Schwarzschild's solution for a static, spherically symmetric point mass [60, 61] to allow for the possibility of non-zero Λ. Kottler's solution,

$$ds^2 = -\left(1 - \frac{2M}{r} - \frac{\Lambda}{3}r^2\right)c^2dt^2 + \frac{dr^2}{1 - \frac{2M}{r} - \frac{\Lambda}{3}r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{1.6}$$

where M is called the *mass parameter*, reduces to a representation either of Schwarzschild's or de Sitter's geometry when $\Lambda = 0$ or $M = 0$, respectively, as Eddington had finally shown in 1923 [1]; thus, the geometry Kottler had first described generally is known as the Schwarzschild-de Sitter (SdS) geometry, and we refer to Eq. (1.6) as the statical SdS metric. It will be central to our analysis in Chapter 4, through which it will be shown to describe a universe that would theoretically appear as ours does.

Now, from Eq. (1.5) Eddington found that the redshift de Sitter had predicted as a phenomenon of his static geometry was in fact due to a repulsion brought in by Λ [1]; viz., he showed that particles could not remain at constant physical distances from each other in the geometry described by Eq. (1.5), but would recede according to

$$\frac{d^2r}{ds^2} = \frac{\Lambda}{3}r, \tag{1.7}$$

so that while every particle would claim to remain at rest (at $r = 0$) in de Sitter space, each would say that all others are exponentially receding. Then, due to the implicit symmetry of the geometry—i.e. from Eq. (1.3)—Eddington concluded that all observers in de Sitter's geometry would observe this same phenomenon, and independently arrived at an expanding solution by 1923. This he substantiated with an updated list of recession velocities of the spiral nebulae provided by Slipher, which gave 36 : 41 odds in favour of the expansion.

Eddington argued that the singularity de Sitter had found at $\chi = \pi/2$—a 'paradox' which de Sitter had dismissed, as Eddington put it, because its effects would be relevant only to events 'before the beginning or after the end of eternity' [1]—could not be physical; it is coordinate-dependent, existing at a radius $r = R = \sqrt{3/\Lambda}$ from every observer in the statical solution, Eq. (1.5), and thus the entire sphere can be covered by intersections of the 'lunes' of neighbouring particles, though they could never share information received from outside each other's respective 'lune'.

Eddington's argument sufficiently removed the paradox of the de Sitter horizon, but his solution still contained the coordinate singularity, and he had even suggested that this singularity could not be removed in a meaningful coordinate basis [1]. This would eventually be proven false: the first coordinate transformation which removed the singularity was found by Georges Lemaître in 1925, while he studied at MIT [42]. Eddington had noticed, in

[1], that de Sitter's original solution, Eq. (1.4), did not contain the proper time coordinate, and was therefore not a good representation of the geometry, and proposed that the static line-element, Eq. (1.5), used the correct basis. However, what Lemaître noticed was that in de Sitter's solution constant spatial coordinates are not geodesics—de Sitter's frame is not *comoving*. Instead, he noted that de Sitter's system has a preferred location at $\chi = 0$. This is equally true of the origin in the static line-element proposed by Eddington. What Lemaître discovered instead, was that a reference frame which treats all points in space on equal footing—a truly comoving system of geodesics in de Sitter space—is an exponentially expanding Galilean system, in which the coordinate singularity occurs at $t = \pm\infty$. In modern notation, this solution is written

$$ds^2 = -dt^2 + e^{2Ht}[dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)], \tag{1.8}$$

where $H \equiv 1/R = \sqrt{\Lambda/3}$, and we see that the geometry is both homogeneous and isotropic, as in Einstein's model. Lemaître would conclude [42],

> Our treatment evidences [the] non-statical character of de Sitter's world which gives a possible interpretation of the mean receding motion of spiral nebulae.
> ... [However,] we are led back to the euclidean space and to the impossibility of filling up an infinite space with matter which cannot but be finite. De Sitter's solution has to be abandoned, not because it is non-static, but because it does not give a finite space without introducing an impossible boundary.

This represents an important step in the progression of Lemaître's understanding, which unfortunately went largely unnoticed by cosmologists at the time—his paper having been largely unknown. Due to Einstein's and Eddington's influence, Lemaître, among others, had believed very strongly in the concept of a closed, unbounded spherical Universe, which he forever retained in his cosmological models. In the publication of his static model in 1917 [56], Einstein had noticed that in order to describe the Universe it would be incorrect to assume, as he did in his calculation of the perihelion precession of Mercury, that space would be asymptotically flat. And the ingenuity with which Einstein solved this problem, by introducing his spherical universe, was noted later by Eddington [2]:

> I think Einstein showed his greatness in the simple and drastic way in which he disposed of difficulties at infinity. He abolished infinity. He slightly altered his equations so as to make space at great distances bend round until it closed up. So that, if in Einstein's space you keep going right on in one direction you do not get to infinity; you find yourself back at your starting-point again. Since there was no longer an infinity, there could be no difficulties at infinity. Q.E.D.

Eddington was convinced already by 1923 [1] that Λ is a Universal constant which gives the radius of curvature of the vacuum at every point, as $\sqrt{3/\Lambda}$, and that 'The length of a specified material structure bears a constant ratio to the radius of curvature of the world at the place and in the direction in which it lies'; thus, he associated the fact that material objects are measured to have the same lengths, regardless of location, with the positive constancy of Λ. He summarised his argument,

> From this point of view it is inevitable that the constant Λ cannot be zero; so that empty space has a finite radius of curvature relative to familiar standards. An electron could never decide how large it ought to be unless there existed some length independent of itself for it to compare itself with.

Lemaître was well aware of, and much influenced by both of these theories, and therefore sought a universal model which would account for both, as Einstein's model did, but which would also account for the redshifts of the nebulae, which only an 'impossibly bounded' de Sitter model could do. He therefore sought 'an intermediate solution which could combine the advantages of both' the Einstein and de Sitter universes [62]. This, he achieved in 1927 [63], in a paper which would unfortunately remain unnoticed until after he brought it to Eddington's attention a few years later, following Hubble's discovery.

But Lemaître's solution was in fact a rediscovery of an earlier solution found by Aleksandr Friedman in 1922, which was unknown to most western astronomers and theoreticians until after Hubble's discovery. In two separate papers [64, 65, 66, 67], Friedman published models for isotropic, homogeneous, expanding/contracting matter-filled universes with constant positive and negative curvature, whose dynamical equations (written in modern notation),

$$3\frac{\dot{a}^2 + K}{a^2} - \Lambda = \kappa\rho \tag{1.9}$$

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2 + K}{a^2} - \Lambda = -\kappa p, \tag{1.10}$$

where $\kappa = 8\pi G/c^4$ is the proportionality constant from the Einstein equation, Eq. (1.2), describe a time-dependent scale factor, $a = a(t)$, typically scaled by requiring the value of the curvature constant, K, to be -1 or $+1$. Note that Friedman did not explicitly treat the Euclidean case, for which $K = 0$; nor did he account for non-zero pressure. The solution Lemaître had rediscovered in [63] was the case $K = +1$, with generalisation to $p \neq 0$, in accordance with his understanding of the Universe. He further analysed the astronomical aspects of this solution, providing a description of the receding velocities of the nebulae according to the expansion of space [68].

The pressureless, dust-filled Einstein model may be recovered from Eqs. (1.9) and (1.10) by requiring $\ddot{a} = \dot{a} = 0$, $\rho = \rho_m$, and $p = p_m = 0$, for which we obtain the result,

$$\Lambda_E = \frac{\kappa}{2}\rho_m = K/a_E^2 = +1/a_E^2. \tag{1.11}$$

Lemaître's comoving de Sitter geometry, Eq. (1.8), is equivalent to Friedman's solution with $a(t) = e^{Ht}$ and $p = \rho = K = 0$. All are special cases of the general form of the Robertson-Walker metrics, Eq. (1.12), discussed at the end of this section.

As mentioned above, the works of Friedman and Lemaître went effectively unnoticed throughout the 1920s. Einstein had known of Friedman's original article, but at first thought Friedman had made a calculation error. In a note he submitted shortly after its publication [69], he claimed that Friedman's result seemed suspicious ('verdächtig'); however, in a retraction he submitted the following year [70], after Friedman had managed to deliver him an explanation [68], he called Friedman's result correct and clarifying ('richtig und aufklärend').

In spite of this, Einstein would make no further reference to Friedman's work until after Hubble had sufficiently demonstrated the Universal expansion; indeed, the Einstein and de Sitter universes remained the two contending geometries of cosmological theories until 1930. For instance, in 1926, following the publication of his famous result for the distance to the Andromeda nebula, Hubble began considering the possibility of a static, homogeneous spatial distribution of galaxies, and from his result estimated the radius of curvature for the Einstein Universe, $a_E = 2.7 \times 10^{10}$ pc [54]. Then, in 1929, he made note of the de Sitter expansion as the possible cause of the linear redshift-distance relation [55], which was a known first-order approximation of the redshift in de Sitter's geometry—the relevant calculation of this, by Howard Robertson in 1928 [43], which contains references to the earlier work on this subject, is discussed below.

Then, in 1930 Lemaître made known to Eddington his previous result, [63]. In his famous paper in which he proved Einstein's static universe was unstable [71], Eddington begins:

> Working in conjunction with Mr. G. C. McVittie, I began some months ago to examine whether Einstein's spherical universe is stable. Before our investigation was complete we learnt of a paper by Abbé G. Lemaître which gives a remarkably complete solution of the various questions connected with the Einstein and de Sitter cosmogonies.

Then, in the same month as Eddington's article had appeared (May), de Sitter published an article in which Friedman is credited as the original discoverer of the expanding universe solution [72]:

> A dynamical solution . . . is given by Dr. G. Lemaître in a paper published in 1927, which had unfortunately escaped my notice until my attention was called to it by Professor Eddington a few weeks ago.

In de Sitter's citation to Lemaître's article [63], he adds a note: 'A similar solution had previously been given by A. Friedman . . . ', citing [64].

Following this exposition of the previously unnoticed works on the expansion paradigm, there was a flurry of articles published which would model the expanding Universe. In March 1931, Lemaître published an (updated) English translation of his 1927 article [62], along with a further analysis of his considerations of the beginning of expansion from a dense Einstein universe [73], which elaborated on Eddington's previous analysis [71]. In this paper, Lemaître generalised the proof of the Jebsen [36, 37]-Birkhoff [38] theorem (see [74]) to non-zero Λ, showing that the geometry outside a (nonrotating and uncharged) spherical mass is SdS, even if expanding or contracting. Thus, Lemaître proved that the formation of condensations in an Einstein universe would not lead directly to expansion or contraction of the universe [73]. Then in May of the same year, Lemaître submitted his proposal for 'The Beginning of the World from the Point of View of Quantum Theory' [75]; from a singularity which Fred Hoyle would later dub the 'Big Bang'. Lemaître's later work, e.g. [76, 77, 78], would be aimed at describing the present Universe as having long since expanded beyond its Einstein radius, where it had remained for some time in a nearly static state due to only a very slight imbalance in favour of the cosmic repulsion of the Λ-term over the Newtonian attraction.

However, in April 1931, Einstein published a paper which would take a markedly different approach. In his now-famous retraction of the Λ-term [79], he suggested a cycloidal model[1] for the Universe—the well-known 'closed Einstein-de Sitter model'—which was essentially a re-statement of Friedman's original solution for $\Lambda = 0$, substantiated by Hubble's discovery. A few months later, in July, Otto Heckmann published the first suggestion within the expanding Universe paradigm that Λ and the spatial curvature K might also be negative or zero [82].

Note, that the solutions studied by Heckmann, on which he elaborated in a second article published the following February [83]—e.g., plotting all solutions for the nine general models in which relativistic and non-relativistic matter do not interact,—had been published by Robertson already in October 1929 [84], well after Hubble had sufficiently demonstrated the linear velocity-distance relation of the galaxies [55]. We must therefore dwell briefly on this paper by Robertson, because it is in fact very unclear, from it alone, which stance Robertson had taken on the matter of expansion—although it has been remarked that he had 'discarded [the expanding models] in favour of static universes' [68],—and because there is a direct link to the proposal of one of the most important aspects of all cosmological theories, known as 'Weyl's postulate' [24, 25, 27].

The successful goal of [84] was to derive, in general, the isotropic and homogeneous geometries. However, Robertson failed to describe the redshift by formulating the evolution of the scale factor in a meaningful way—e.g., by relating his results to Friedman's equations. The paper spends little effort in describing the non-stationary solutions, aside from including them in general formulations. Instead, Robertson always makes special effort to relate his results to a second assumption, that the Universe be stationary. From these assumptions alone, he proved that the only possible stationary cosmologies (with positive curvature) are those of Einstein and de Sitter. Among these, he shows that only the de Sitter cosmology can account for the observed redshifts. He seems not to have taken preference to any of the solutions he had derived, but had kept to his goal of an analytical proof that geometries which had previously been known, but had previously only been given as solutions which would satisfy the field equations in a specific way, rather than as special cases of a more general solution, are the only ones that exist.

However, his tendency to give the stationary solutions special treatment, along with the result that Einstein's universe contains no systematic redshift, is suggestive of a preference towards the de Sitter solution over Friedman's and Einstein's. This is evident in the work he had then been carrying out in the same vein as Hermann Weyl. In 1923, Weyl published an analysis of de Sitter's geometry—first, in the fifth appendix to his book *Raum, Zeit, Materie* (*Space, Time, Matter*) [85], which was never translated to English, and then in a

[1] Note that in [79] Einstein concentrates only on one period of the cycloid to describe our Universe, and makes no mention of the possibility of a periodic behaviour, as in 'big bounce' or 'phoenix' theories. The periodic behaviour of this universe was in fact first suggested by Friedman in [64, 65], although Einstein's analysis in [79] had perhaps made this more plain. However, credit for the first serious proposal of an oscillating universe must be given to Richard Tolman, who presented a thermodynamic model of the sort in 1931 [80]. In this paper, Tolman writes, 'Recently, . . . a simple model of the universe has been discussed by Einstein [79] which exhibits a possibility for a quasi-periodic solution of a type which must now also be considered.' It is likely the ambiguity of this statement which has led some—e.g., [81]—to credit Einstein with the suggestion; for one can easily mis-interpret from what Tolman wrote, that Einstein had made note of the possibility of quasi-periodicity. In fact, Tolman's reference to Einstein's work ends at the citation.

supplementary paper [20, 21]; see the editor's note, [26], to the reprint of [20]—which was unique from Eddington's approach in that it described an expanding kinematical system, from which two key advancements were made: the first calculation of redshift in a non-static solution (which, to first order, anticipated the linear redshift-distance relation found later by Hubble [55]), and the proposal of a coherency of matter which describes the concept of a cosmic standard time, the importance of which was noted in particular by Robertson [84, 24], who had referred to it as 'Weyl's postulate' in [24], and then later exposed by Hermann Bondi [25], who is credited with actually giving it this name. On this postulate, Weyl wrote that the assumption that two stars would 'belong to the same system and [be] causally connected through a common origin, implies that their world lines have the same action domain Σ.*' [21]. Weyl's model defined the universe as the three-dimensional slice of de Sitter spacetime that Lemaître later described with his line-element, Eq. (1.8), which would cover half of the entire 'sphere', bounded by the infinitely distant past and infinitely distant future—c.f. the singularity at $\chi = \pi/2$ in de Sitter's coordinates, Eq. (1.4),—of which any observer would see only a 'cuneiform sector' [23]. This universe was to have begun an infinite time in the past, when the distances between stars would tend to zero, and would expand according to the cosmic repulsion described by Λ [21]:

> All stars of our system Σ flee from any arbitrary star in radial directions; there is inherent in matter a universal tendency to expand which finds its expression in the "cosmological term" of EINSTEIN'S law of gravitation . . . .

In 1928, Robertson proposed a formally equivalent model to Weyl's [43], in which he operated from comoving coordinates derived independently from Lemaître. In a first-order calculation, he derived a linear velocity-distance relation, to which he noted Weyl's result would also reduce. This cosmological theory was tidied up in a paper written by Weyl which was submitted in July 1929 [23], after he had visited both Robertson and Tolman in the United States. He proposed that the Universe began from a common origin at $t = -\infty$ and subsequently expanded as a homogeneous matter distribution of infinitely small density with open space and infinite mass, and updated his derivation of the observed redshift using the comoving coordinates. It was likely this theory which Robertson had in the back of his mind when he wrote [84], which may have been the reason that in it he paid little attention to the non-static solutions. However, the theory never became popular, probably because [23] was published in May 1930—the same month as Eddington's and de Sitter's papers on the expanding Universe, which exposed Lemaître's previous work.

Thus, although it seems that Robertson had favoured the de Sitter solution in [84], this cannot have been because he had hoped it would describe a non-expanding picture of the Universe. Hubble's demonstration of the linear velocity-distance relation of extra-galactic nebulae had been published eight months prior, and had actually been anticipated by Robertson, among others. Instead, there seems to be an implicit assumption in [84], that the true physical cosmology must have a stationary geometrical representation. This bias is not explicitly voiced, but it seems reasonable to speculate that it did exist, and that it was related

*Both the papers by de Sitter [57] and Eddington [1] lack this assumption on the "state of rest" of stars—by the way the only possible one compatible with the homogeneity of space and time. Without such an assumption nothing can be known about the redshift, of course. [Weyl's footnote.]

to one of the cosmological conditions that Tolman had suggested in April that year [86]; viz. that it be possible to write the line element 'in a form which is spherically symmetrical in the spatial variables, symmetrical with respect to past and future time, and static with respect to time.' Tolman's physical motivation for proposing such a condition was not that he had thought the Universe was properly static, but that, respectively, each element of the condition would require a universe which, on the large scale, did not have different properties in different directions, would have large-scale reversible behaviour, and would, by-and-large, be in a steady state [86]. These requirements would be rejected within a year, when it was understood that an expanding, physical Universe could be realised without them.

Shortly thereafter, came Heckmann's popular realisation that expanding universes can be described with all values of K and Λ. In particular, my previous statement that Heckmann's was the first treatment of the entire gamut of homogeneous, isotropic models within the expanding Universe paradigm is now justified; for though Robertson had found these solutions prior to Heckmann's paper, [82], he did not attempt to expose them as potential cosmological models. Heckmann had thus been the first to understand that both K and Λ might be zero in the expanding Universe, as one of nine general classifications of these geometries. Thus, his papers prompted Einstein and de Sitter, in March 1932, to propose an idealised dust model for the Universe which completely neglects Λ, and would further neglect K as a first approximation [87]. This is known as the Einstein-de Sitter model, and was popularly thought to approximate the correct Universal geometry during the remainder of the twentieth century.

We should wonder why Einstein and de Sitter did not give credit to Friedman, rather than Heckmann, for this suggestion, because the former had considered general Λ, and the case of zero curvature is a simple limit of his solutions. The important leap which they credited to Heckmann, was that he had pointed out 'that the non-static solutions of the field equations of the general theory of relativity with constant density do not necessarily imply a positive curvature of three-dimensional space, but that this curvature may also be negative or zero.' But Friedman had found the negative curvature solution already in 1924, so this is obviously not true. In fact, neither Einstein nor de Sitter had been aware of Friedman's second paper, [66, 67]. In 1933 [88], de Sitter wrote,

> Friedmann discusses the solutions of the field equations for different values of Λ. Lemaître considers only positive Λ. Both authors consider a positive curvature of space only. The fact that both Λ and the curvature may as well be negative or zero was only pointed out by Dr. Heckmann in July 1931.

The book from which this is taken, *The Astronomical Aspect of the Theory of Relativity*, was published only one year before de Sitter's death, so he likely never knew of Friedman's second paper; but Einstein may also never have known of it. In his writings on the cosmological problem following the paradigm shift in 1929 – 1930—e.g., in [79], as well as in his treatment of the problem in an appendix to *The Meaning of Relativity* which he had published in 1945 [89]—Einstein only references Friedman's first paper [64, 65] when crediting him with the original analytical discovery of the expanding Universe.

The situation is yet more confusing when we note that Robertson had known of both papers by Friedman in 1929 [84], when he had favoured de Sitter's universe, and before the first one had been publicised by de Sitter; and that in [88], in the paragraph *directly preceding* the one which contains the above quotation, de Sitter actually references [84] in connection with the proof that the Einstein and de Sitter universes 'are the only possible static, homogeneous, and isotropic solutions with positive curvature'. A further level of confusion comes in a review on the various descriptions of cosmology, written by Robertson in 1933 [24], in which he notes that de Sitter did write an article in 1932 [90], where he performed an 'analysis and classification of all of Friedman's worlds $p = 0$.' However, in this article de Sitter again only references Friedman's original paper [64, 65], citing Heckmann's paper [82] as the first suggestion of the nine possible expanding worlds. Although Robertson's statement is technically correct, it is misleading because he had not noticed that de Sitter had been unaware of the second paper, [66, 67].

The prevalent concept in all of the proposed descriptions of the Universe discussed in this section, is that of a separation of spacetime into homogeneous, isotropic spatial slices, comoving in proper time. This is due to the observation that the Universe appears to us to be isotropic, and to the cosmological principle—the idea that any observer in any galaxy must make essentially the same observations of the cosmos. The further assumption, that such a time exists, is what Bondi called 'Weyl's postulate' [25], although this is more restrictive than Weyl's original proposal, which had to do with a more general notion of causal coherence [26]. Working from essentially these assumptions, Arthur Walker [91, 92, 93] and Robertson [94, 95, 96] rigorously proved that the only kinematical systems which can exist can be represented by the three classes of isotropic, homogeneous spacetime geometries,

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \right], \tag{1.12}$$

for which the factor $a(t)$ is conventionally scaled so that the curvature K can be either $+1$, 0, or -1; i.e. the three non-static solutions Robertson had found in [84]. These metrics are therefore called the Robertson-Walker (RW) metrics, and the cosmological models they describe are called Friedman-Lemaître-Robertson-Walker (FLRW) models for the relevant contributions all four made to the development of the theory.

On the above assumptions, Robertson wrote, in 1929 [84],

> ... we have described the actual world in terms of coördinates which effect a natural separation of it into space and time and have determined its ideal background by a single assumption ... which is but the concrete expression of a uniformity implied by the very concept of a system of relativistic cosmology; in requiring that this ideal background can be fitted into the actual universe in the way described we have expressed in another way the assumption made by other writers on the subject, above all H. Weyl, that the world lines of all matter in the universe form a coherent pencil of geodesics.

Thus, Robertson had realised that this 'natural separation' of space and time is not exactly what Weyl had intended. But regardless, this assumption, which may have been first explicitly stated by Robertson, and *is* anyhow related to Weyl's postulate, has remained an elementary facet of standard relativistic cosmology, as a multitude of FLRW models have been used to describe popular cosmological theories ever since. As such, the general formulation of the cosmological problem has commonly been to define, at the outset, this division of spacetime into space and one Universal time, applicable both as the proper time of the Universe as a whole, and as that of every cosmic rest-frame observer, simultaneously; then, due to astronomical observations and the cosmological principle, to assume these spatial sections must be homogeneous and isotropic.

In § 4.3, a cosmological model is described which satisfies the cosmological principle, observed isotropy, and, in a more general sense, Weyl's prostulate, but which is geometrically dissimilar from the RW spacetimes, and thus disproves the common assertion that these geometries alone satisfy the principles of relativistic cosmology; and it serves to illustrate the error of Robertson's 'natural separation' of spacetime, as, in that case, the cosmic standard time corresponding to Weyl's 'assumption on the "state of rest" of stars' [21], is clearly not a synchronous attribute of the proper times of each of those stars. However, the reasonable advantage of this model cannot be appreciated without significant theoretical development; so, we continue our description of the development of cosmological theory, in which we will rediscover the more natural and rational interpretation of cosmic expansion which was initially expounded by Eddington, which is formally inconsistent with FLRW big bang cosmology.

1.2 Interpreting Cosmic Expansion

Following Hubble's observational results, there emerged two schools of thought regarding Λ, and the one that won out can be generally associated with the notion, that 'Λ is *not necessary* as long as the Universe is not static'. Although many words were used over the years, becoming more severe as time went on and the paradigm solidified, 'unnecessary' seems to be the best to describe this interpretation, and was anyhow the original idea for which it became accepted. This interpretation is called the Einstein interpretation, for although it was first suggested by Friedman [64, 65], then Heckmann [82], and was favoured by many others, it had its strongest support from Einstein, who completely rejected the Λ-term and strongly promoted the $\Lambda \equiv 0$ philosophy throughout the rest of his life. Many authors, e.g. [58, 97, 98, 99], have written accounts of Einstein's interpretation of Λ, but it is worthwhile to produce a full exposition of this idea here because many of the others are either brief, misleading, or even incorrect.

Scepticism towards Einstein's introduction of the Λ-term in 1917 [56] came almost immediately. When de Sitter proposed his solution later that year [57], he wrote,

> It cannot be denied that the introduction of the constant $\Lambda\ldots$ is somewhat artificial, and detracts from the simplicity and elegance of the original theory of 1915, one of whose great charms was that it embraced so much without introducing any new empirical constant.

As an astronomer, de Sitter had however been confronted with a necessity to account for the redshifts of spiral nebulae throughout the 1920s; thus, his own sentiments towards the Λ-term had become less severe by the time of Hubble's discovery, due to Eddington's interpretation of a cosmic dispersal in a de Sitter cosmology. In the majority of de Sitter's articles from 1930 until his death in 1934 (see, e.g., [72, 88, 100]), with the exception of the one he co-authored with Einstein in 1932 [87], he had invoked the Λ-term to describe the present nature of the expansion inferred from Hubble's results. But although de Sitter's initial reaction towards Λ had become less severe due to Eddington's and Hubble's results, the influence of such sentiments as the one he initially expressed would be everlasting on Einstein, who, in his later writings on the cosmological problem, would emulate de Sitter's initial misgivings.

Einstein's distaste for the Λ-term is already evidenced in a postcard he sent to Weyl on May 23, 1923 [Einstein Archives call number [24-81.00]; original held by the Weyl-Archiv at the Eidgenössische Technische Hochschule Zürich], following the publication of Weyl's cosmological theory. In this letter, he wrote,

> [Des kosmologischen Gliedes] würde zunächst nichts schaden, weil man die Theorie ganz zwanglos verallgemeinern kann, was ich unterdessen gefunden habe.. . . Ich sende Ihnen dann die Korrektur meiner Arbeit, die ich unbedingt publizieren muss, weil der Eddington'sche Gedanke notwendig zu Ende gedacht werden muss. Ich glaube jetzt auch, dass alle diese Versuche auf rein formaler Basis die physikalische Erkenntnis nicht weiter bringen werden. Vielleicht hat die Feldtheorie schon alles hergegeben, was in ihren Möglichkeiten liegt. Inbezug auf das kosmologische Problem bin ich nicht Ihrer Meinung. Nach de Sitter laufen zwei genügend von einander entfernte materielle Punkte beschleunigt auseinander. Wenn schon keine quasi-statistische Welt, dann fort mit dem kosmologischen Glied. [[The cosmological term] would not hurt at first, because one can generalise the theory quite naturally, as I have meanwhile found.. . . I send you then the correction of my work, which I absolutely must publish because the Eddington-ish considerations must necessarily be thought through. I now also believe that all these attempts towards a purely formal basis will not lead to further physical knowledge. Perhaps the field theory has already given away everything that lies in its possibilities. With reference to the cosmological problem, I am not of your opinion. Following de Sitter, we know that two sufficiently separate material points are accelerated from one another. If there is no quasi-static world, then away with the cosmological term.]

Eight days later, Einstein submitted his second note on Friedman's paper [70], which he now referred to as 'correct and clarifying'. His use of the word clarifying ('aufklärend') in the context of a retraction of his previous statement should have seemed puzzling without this letter to Weyl; however, now that we have it in the context of the other interpretations of expansion in de Sitter spacetime by Eddington and Weyl, it is clear that Einstein had been expressing his preference for Friedman's solution as a possible way out of the alternate description of accelerating expansion.

But Friedman's solution had not been known to everyone, and Weyl's confusion over this letter is quite plain in a dialogue he wrote the following year, on inertial mass and the cosmos, in which an orthodox 'Petrus' (St. Peter), who is supposed to represent Einstein's position, eventually responds to a more radical—in fact, as Weyl must have seen things, dangerously close to apostatical[2]—'Paulus' (St. Paul), who presents Weyl's ideas [22],

[2] At a point early on in the dialogue, 'Paulus' actually says, 'wenn hier der Fels liegt, auf dem die Relativitätskirche steht, o Petrus!, so bin ich in der Tat ein Abtrünniger geworden. Aber um dich über meine Ketzerei ein wenig zu beruhigen, . . . ' (if here lies the rock on which the relativity church stands [cf. Matthew 16:18], o Peter!, then I have indeed become an apostate. But to calm you a little about my heresy, . . .) [22].

> Wenn es mit Hilfe des kosmologischen Gliedes nicht gelingt, das Machsche Prinzip durchzuführen, so halte ich es überhaupt für zwecklos und bin für die Rückkehr zur elementaren Kosmologie. [If the cosmological term fails to help with leading through to Machs principle, then I consider it to be generally useless, and am for the return to the elementary cosmology.]

'Petrus' was speaking here of a return to a special relativity-based cosmology; and 'Paulus' went on to reason that this would be too hasty, explaining that the expanding 'de Sitter cosmology' would account better for the observational data than either this 'elementary cosmology' or Einstein's statical model. Weyl obviously was, like nearly everyone else in 1924, completely unaware of Friedman's original solution, and it would seem oddly devious if Einstein, who had apparently been pleased with the clarification that had been afforded by Friedman's solution, actually made no subsequent attempt to clear up this matter by sharing what he knew.

But regardless of the way circumstances played out, it is clear that Einstein had been unsatisfied with the Λ-term throughout most of the 1920s, and it is therefore not surprising that when he submitted his paper on the cosmological problem in 1931 [79] he would attempt to retract it. He mentions the Λ-term with negative connotations twice in this article, from which we see hints of de Sitter's speculation. After giving an introduction to the new picture of the expanding Universe, Einstein writes,

> Unter diesen Umständen [FRIEDMANNS und HUBBELS Resultate] muß man sich die Frage vorlegen, ob man den Tatsachen ohne die einführung des theoretisch ohnedies unbefriedigenden Λ-Gliedes gerecht werden kann. [With these understandings [Friedman's and Hubble's results], one must ask oneself whether the facts can be accounted for without the introduction of the (theoretically, anyhow) unsatisfactory Λ-term.]

Then, in conclusion,

> ... Bemerkenswert ist vor allem, daß die allgemeine Relativitätstheorie HUBBELS neuen Tatsachen ungezwungener (nämlich ohne Λ-Glied) gerecht werden zu können-scheint als dem nun empirisch in die Ferne gerückten Postulat von der quasistatischen Natur des Raumes. [Above all, it is remarkable that Hubble's new facts allow general relativity theory to seem less contrived (namely, without the Λ-term), as the postulate of the quasistatic nature of space has moved into the distance.]

Thus, Einstein had thought of the Λ-term as a blemish to his theory, which had been added *ad hoc*, in order to account for an incorrect world-view. With these remarks, he seems glad to be rid of the term which had not only itself seemed contrived, but which he feared had made general relativity theory seem contrived, allowing its critics an argument which could now be revoked.

A year later, Einstein and de Sitter published their cosmological model [87], in which Λ was to be neglected from the outset. The treatment of Λ in this article was mis-interpreted by Subrahmanyan Chandrasekhar [97], who wrote,

> ... When Friedmann's cosmological models were found to provide an adequate base for accounting for the simple fact of the Hubble expansion, Einstein and

de Sitter, in a joint paper, stated that one can do without the cosmical constant. In view of the many exaggerated statements that have been made concerning this supposed 'retraction' of Λ, it is of interest to record precisely what it was they said.

> Historically the term containing the "cosmological constant" Λ was introduced into the field equations in order to enable us to account theoretically for the existence of a finite mean density in a static universe. It now appears that in the dynamical case this end can be reached without the introduction of Λ...
>
> ... The curvature [constant Λ] is, however, essentially determinable, and an increase in the precision of data derived from observations will enable us in the future to fix its sign and to determine its value.

The edit in the second paragraph was added by Chandrasekhar in order to support his argument that Einstein and de Sitter had not rejected Λ. However, it is absolutely clear in the two-page article [87], that the curvature of which they speak is not Λ, but the curvature of three-dimensional space, K; and it is shocking that Chandrasekhar did not understand this. Regardless, the cosmological constant was neglected altogether in the Einstein-de Sitter model because in the full FLRW cosmological models Λ truly is unnecessary in order to account for the observed expansion, as Friedman and Heckmann had previously noted.

So, as of 1932, Einstein had (jointly) proposed a model of the Universe which rejected the Λ-term, which he had previously thought to be theoretically unsatisfactory. These thoughts appear to have strengthened in subsequent years, and in 1945 he added an appendix to *The Meaning of Relativity* [89], in which he includes many damning remarks towards the existence of Λ; for example,

> ... The introduction of [Λ] constitutes a complication of [general relativity] theory, which seriously reduces its logical simplicity. Its introduction can only be justified by the difficulty produced by the almost unavoidable introduction of a finite average density of matter. We may remark, by the way, that in Newton's theory there exists the same difficulty.
>
> The mathematician Friedman found a way out of this dilemma.* His result then found a surprising confirmation by Hubble's discovery of the expansion of the stellar system.

And then, in a summary note,

> The introduction of the "cosmologic member" into the equations of gravity, though possible from the point of view of relativity, is to be rejected from the point of view of logical economy. As Friedman was the first to show one can reconcile an everywhere finite density of matter with the original form of the

*He showed that it is possible, according to the field equations, to have a finite density in the whole (three-dimensional) space, without enlarging the field equations *ad hoc*. Zeitschr. f. Phys. 10 (1922). [Einstein's footnote.]

> equations of gravity if one admits the time variability of the metric distance of two mass points.*

Eventually, Einstein's overall interpretation of Λ was appropriately summarised by Wolfgang Pauli in his 1958 supplemental notes to the English addition of *Theory of Relativity* [101], which had originally been published in German in 1921:

> Einstein was soon aware of [the results of Friedman, Lemaître, and Hubble] and *completely rejected the cosmological term* as superfluous and no longer justified. I fully accept this new standpoint of Einstein's.

Finally, in 1970, when George Gamow's autobiography was published posthumously, we learned that 'much later [than Hubble's discovery], when [Gamow] was discussing cosmological problems with Einstein, he remarked that the introduction of the cosmological term was the biggest blunder he ever made in his life' [102]. Now, whether this was considered a blunder because he may otherwise have discovered Friedman's solution and predicted Hubble's result; or because now that he had suggested the term it would never go away completely, cropping up, e.g., as an integral component in Eddington's and Lemaître's models, and later in the steady state theory; or because of something else entirely, we'll never know. In fact, we can never truly know whether these *were* his words because they came second hand, as well as posthumously on both accounts!

However the true statement may have been made or meant, this completes our comprehensive exposition of Einstein's interpretation of Λ in light of the observed cosmic expansion. In his own words, Einstein did abandon the cosmological term unequivocally, and Pauli's summary of Einstein's view seems appropriate. Although the description of this view was given gradually, over many years and with increasingly severe tones, it always remained mathematically equivalent to the Einstein-de Sitter model, proposed in 1932 [87]. And as may have been expected when the master of the theory had cast his lot, the *status quo* went with him. Thus, the Einstein-de Sitter model, effectively parametrised by dust-density and curvature, became the standard model in cosmology throughout most of the twentieth century—and it was for many a great shock to find that the Universe is otherwise.

This should not have been so; for there is a natural explanation of expansion which was popularly neglected, although many of the prominent researchers in early cosmology—most notably, Eddington—had tried to make it work. This interpretation, which I have briefly mentioned above, is that the cosmic expansion is driven—and in fact has been for all time—by positive Λ. There are difficulties with this 'Eddington interpretation' which have always been known; most notably, that, with the exception of Weyl's model, it cannot be accounted for in any FLRW big bang universe. For as we follow Friedman's equations, Eqs. (1.9) – (1.10), backwards in time, the constant Λ must inevitably become negligible when compared with the increasing density of matter. Accordingly, the Einstein interpretation is the only interpretation of cosmic expansion that has existed in the literature since the discovery of

*If Hubble's expansion had been discovered at the time of the creation of the general theory of relativity, the cosmologic member would never have been introduced. It seems now so much less justified to introduce such a member into the field equations, since its introduction loses its sole original justification,—that of leading to a natural solution of the cosmologic problem. [Einstein's footnote] the cosmic microwave background (CMB) in 1965, which provided strong confirmation of the Big Bang theory.

Therefore, the overwhelming evidence for positive Λ, demonstrated from observations of type Ia supernovae (SNe Ia) by two separate groups at the end of the last century [49, 50], allowed for only a superficial contrast between the interpretations of cosmic expansion and observational results. Furthermore, quantum theory had already predicted a non-zero vacuum energy which could act like the observed cosmological constant at late times, after having evolved from a much larger value early on. Since the discovery of non-zero Λ, there has been much hope for statistical evidence of such a *quintessence* field, which would decay in such a manner, and could thus be attributed as a solution to the *cosmological constant problem*—the discrepancy between values of the vacuum energy predicted by quantum field theory and observed Λ—and a suggestive link to inflation theory; see, e.g., [103] for a recent comprehensive treatment of this picture. This prediction, along with the fact that Λ has no essential role in the standard cosmology, resulted, almost immediately after its discovery, in re-naming the observed cosmological constant, rather as 'dark energy',—a blanket term to cover all theories.

In the remainder of this chapter, I will continue the historical account of cosmology and the cosmological constant from the Eddington interpretation onwards, and eventually argue against the above theories in that context. The result of this exposition will be that the standard cosmology, which has been increasingly refined since the 1930s, and is now known to be wrought with many problems—e.g. the flatness, horizon, and cosmological constant problems which the quintessence and inflation theories aim to resolve—should likely need to be abandoned in favour of one that would account *a priori* for the facts of observation.

To begin, let us ignore the evidence of the Big Bang—which will anyhow be shown to have a description in the new cosmology of Chapter 4—so that the full argument may be presented in an historical context, without having to accept that the new cosmological model, which is consistent with this interpretation, does work. Then we are perfectly at liberty, as Eddington was, to argue the possibility of his interpretation of Λ, which contends that *positive* Λ *must be the sole cause of cosmic expansion.*

Knowing only that the Universe is expanding at a rate which, near to us, looks like a linear velocity-distance relation, the most rational interpretation certainly is that the expansion is driven by positive Λ at all times, rather than as the result of a singular beginning. For the latter can only be described with a full understanding of the process of that beginning, which itself must have been very special in order to account for the uniform expansion we observe everywhere (see, e.g., the discussion of this point in [104, 105]), and anyhow cannot be explained with the FLRW model alone, while the former is a known facet of general relativity theory, which Einstein had to admit was 'possible from the point of view of relativity', and which has the possibility of naturally explaining the observed expansion.

Eddington's belief in the Λ-term did not have the luxury of observational confirmation, as ours now does. He believed Λ was a fundamental constant of the Universe on 'general philosophical grounds', which he had held before and apart from the observed expansion, as mentioned in § 1.1 [1, 13, 106]. But whether or not Eddington's reasons for believing in Λ as a fundamental facet of general relativity theory are ultimately correct, the simple fact remains that Λ is *not* superfluous or unjustified in the mathematical theory, and should not be rejected at the outset. The case $\Lambda = 0$ is very special indeed, and setting $\Lambda \equiv 0$ 'from the point of view of logical economy', so as not to complicate the logical simplicity of the theory, is wrong, and constitutes a breach of the most natural extension of the cosmological principle; i.e. that our *Universe* should not be very special. From this perspective alone, the most logical interpretation of the observation that the Universe is expanding, is that our Universe possesses a positive cosmological constant which drives the expansion; for this is the only interpretation which would go beyond the standard empirical model and provide a natural theoretical *explanation* of it.

In 1933, Eddington published *The Expanding Universe* [2], in which he fully described his interpretation of the cosmic expansion. Although his opinions are often extreme, and will require a little further comment, his idea is made most clear through a string of direct quotations from this book, which concisely represents the picture he held of the Universal expansion:

> The *spiral nebulae* ... are extra-galactic objects; that is to say, they lie beyond the limits of the Milky Way aggregation of stars which is the system to which our sun belongs, and are separated from it by wide gulfs of empty space. ...
>
> Each island system is believed to be an aggregation of thousands of millions of stars with a general resemblance to our own Milky Way system. As in our own system there may be along with the stars great tracts of nebulosity, sometimes luminous, sometimes dark and obscuring. ...
>
> The lesson of humility has so often been brought home to us in astronomy that we almost automatically adopt the view that our own galaxy is not specially distinguished—not more important in the scheme of nature than the millions of other island galaxies. ...
>
> When the collected data as to radial velocities and distances [of these galaxies] are examined a very interesting feature is revealed. The velocities are large, generally very much larger than ordinary stellar velocities. The more distant nebulae have the bigger velocities; the results seem to agree very well with a linear law of increase, the velocity being simply proportional to the distance. The most striking feature is that the galaxies are almost unanimously running away from us. ...
>
> We can exclude the spiral nebulae which are more or less hesitating as to whether they shall leave us by drawing a sphere of rather more than a million light-years radius round our galaxy. *In the region beyond, more than 80 have been observed to be moving outwards, and not one has been found coming in to take their place.*
>
> The inference is that in the course of time all the spiral nebulae will withdraw to a greater distance, evacuating the part of space that we now survey. Ultimately they will be out of reach of our telescopes unless telescopic power is increased to correspond.[3] ...
>
> The unanimity with which the galaxies are running away looks almost as though they had a pointed aversion to us. We wonder why we should be shunned

[3]This speculation has been disproven by Lawrence Krauss and Glenn Starkman [107]. As it turns out, there is no chance of observing light from distant galaxies after a finite time in our future, as the Universe asymptotically approaches de Sitter spacetime.

as though our system were a plague spot in the universe. But that is too hasty an inference, and there is really no reason to think that the animus is especially directed against our galaxy. If this lecture room were to expand to twice its present size, the seats all separating from each other in proportion, you would notice that everyone had moved away from you. Your neighbour who was 2 feet away is now 4 feet away; the man over yonder who was 40 feet away is now 80 feet away. It is not *you* they are avoiding; everyone is having the same experience. ...

...the picture of the universe now in the minds of those who have been engaged in practical exploration of its large-scale features ...is the picture of an *expanding universe*. The super-system of the galaxies is dispersing as a puff of smoke disperses. Sometimes I wonder whether there may not be a greater scale of existence of things, in which it *is* no more than a puff of smoke. ...

...I want to speak of the alteration that Einstein made in his law of gravitation. The amended law is written $G_{\mu\nu} = \Lambda g_{\mu\nu}$, and contains a natural constant Λ called the *cosmical constant*. The term $\Lambda g_{\mu\nu}$ is called the *comsical term*. The constant is so small that in ordinary applications to the solar system, etc., we set it equal to zero, and so revert to the original law $G_{\mu\nu} = 0$. But however small Λ may be, the amended law presents the phenomenon of gravitation to us in a new light, and has greatly helped to an understanding of its real significance; moreover, we have now reason to think that Λ is not so small as to be entirely beyond observation. The nature of the alteration can be stated as follows: the original law stated that a certain geometrical characteristic ($G_{\mu\nu}$) of empty space is always zero; the revised law states that it is always in a constant ratio to another geometrical characteristic ($g_{\mu\nu}$). We may say that the first form of the law utterly dissociates the two characteristics by making one of them zero and therefore independent of the other; the second form intimately connects them. Geometers can invent spaces which have not either of these properties; but actual space, surveyed by physical measurement, is not of so unlimited a nature.

We have already said that the original term in the law gives rise to what is practically the Newtonian attraction between material objects. It is found similarly that the added term ($\Lambda g_{\mu\nu}$) gives rise to a repulsion directly proportional to the distance. Distance from what? Distance from *anywhere*; in particular, distance from the observer. It is a dispersive force like that which I imagined as scattering apart the audience in the lecture-room. Each thinks it is directed away from him. We may say that the repulsion has no centre, or that every point is a centre of repulsion.

Thus in straightening out his law of gravitation to satisfy certain ideal conditions, Einstein almost inadvertently added a repulsive scattering force to the Newtonian attraction of bodies. We call this force the *cosmical repulsion*, for it depends on and is proportional to the cosmical constant. It is utterly imperceptible within the solar system or between the sun and neighbouring stars. But since it increases proportionately to the distance we have only to go far enough to find it appreciable, then strong, and ultimately overwhelming. In practical observation the farthest we have yet gone is 150 million light-years. Well within

> that distance we find that celestial objects are scattering apart as if under a dispersive force. Provisionally we conclude that here cosmical repulsion has become dominant and is responsible for the dispersion. ...
>
> ... [The cosmical constant] renders the theory of gravitation and its relation to space-time measurement so much more illuminating, and indeed self-evident, that to return to the earlier view is unthinkable. I would as soon think of reverting to Newtonian theory as of dropping the cosmical constant. ...
>
> ... There are only two ways of accounting for large receding velocities of the nebulae: (1) They have been produced by an outward directed force as we here suppose, or (2) as large or larger velocities have existed from the beginning of the present order of things.* Several rival explanations of the recession of the nebulae, which do not accept it as evidence of the repulsive force, have been put forward. These necessarily adopt the second alternative, and postulate that the large velocities have existed from the beginning. This might be true; but can scarcely be called an *explanation* of the large velocities. ...
>
> ... the theory recently suggested by Einstein and de Sitter, that in the beginning all the matter created was projected with a radial motion so as to disperse even faster than the present rate of dispersal of galaxies,* leaves me cold. One cannot deny the possibility, but it is difficult to see what mental satisfaction such a theory is supposed to afford. ...
>
> When once it is admitted that there exists everywhere a radius of curvature ready to serve as comparison standard, and that spatial distances are directly or indirectly expressed in terms of this standard, the law of gravitation ($G_{\mu\nu} = \Lambda g_{\mu\nu}$) follows without further assumption; and accordingly the existence of the cosmical constant Λ with the corresponding force of cosmical repulsion is established. Being in this way based on a fundamental necessity of physical space, the position of the cosmical constant seems to me impregnable; and if ever the theory of relativity falls into disrepute the cosmical constant will be the last stronghold to collapse. *To drop the cosmical constant would knock the bottom out of space.* ...

Now, whether Λ should be so *required* in the physical description of gravity by general relativity theory as Eddington argues, is not yet certain. Indeed, Chandrasekhar [97] commented that 'no serious student of relativity is likely to subscribe to Eddington's view that "to set $\Lambda = 0$ is to knock the bottom out of space".' And Eddington's statement, that he would 'as soon think of reverting to Newtonian theory as of dropping the cosmical constant' is obviously exaggerated; for he was intimately aware of the important confirmations the theory had received in describing the perihelion precession of Mercury and the deflection of light,—though he may have been as likely to revert to the theory of modified Newtonian

*For completeness we must add the possible hypothesis that the system once extended much further than now, that it collapsed, and is now on the rebound. This allows the large velocities to have been produced by *inward* directed force, the inward velocities being turned into outward velocities by passage through the centre. So far as I know, this is not advocated by anyone. It does not seem capable of providing for the distribution of velocities which we observe. [Eddington's footnote.]

*They do not state this in words, but it is the meaning of their mathematical formulae. [Eddington's footnote.]

dynamics (MOND) proposed by Mordehai Milgrom [108, 109, 110] as general relativity theory with $\Lambda \equiv 0$, had he been alive in 1983. To Eddington, the theory had only really been complete with the introduction of positive Λ, which, he thought, serves as a physical basis for the length scales the theory would describe. But whether or not anyone would currently subscribe to Eddington's view, to simply dismiss the notions of one of the great natural philosophers of the twentieth century concerning a concept which he had given such considerable thought, without giving the problem an equivalent amount of thought, seems too hasty an act. Thus, I note only that on purely aesthetic or philosophical grounds we can currently make no headway on the matter: for two of the primary exponents of the theory, Einstein and Eddington, had extreme contrasting views. However, we don't need this in order to proceed with Eddington's interpretation of *cosmic expansion*, so it can be left for now as an open issue. All that is required is that $\Lambda \neq 0$ should necessarily be considered a possibility of the mathematical theory of relativity. On this point, no serious student of relativity would currently object, as from a mathematical standpoint Pauli's excessive statements regarding Λ are incorrect!

Working from only this more moderate view and the observation that the Universe is uniformly expanding, Eddington's theory [71], which was anticipated by Lemaître [63, 62], in which the Einstein Universe would expand *due to* Λ was far more natural than the theory that would come to dominate. Eddington's interpretation *explains* the Universal expansion *at all times*, whereas the Einstein interpretation *never* does—for, even if we would set additional conditions in the very early Universe that would contribute to the subsequent physical state of expansion, the inexplicable singularity still remains the initial cause, according to the mathematical solution. However, in order to allow for the implicit expansion of space that is so naturally accounted for by positive Λ, Eddington had to relinquish any possibility of a beginning of expansion, which he regarded anyhow (or perhaps because it did not fit with his interpretation) as 'too unaesthetically abrupt'. As mentioned above, this is true for any FLRW big bang universe aside from Weyl's 'de Sitter cosmology': although expansion may be driven by Λ at late times, it cannot account for the expansion shortly after the big bang, which must be caused by the big bang singularity itself.

1.3 Towards a Standard Cosmogony

In light of the problem Eddington's interpretation has in reconciling itself with the FLRW big bang models, many of the other theorists working on the problem, e.g. Lemaître and de Sitter, would take the more moderate stance, that $\Lambda \neq 0$ is perfectly allowable by general relativity theory, and could possibly have a fundamental role even, so that, despite any uncertainty in that regard, it should always be included in general cosmological descriptions. I've mentioned previously, that, following Hubble's discovery, de Sitter's only apparent deviation from general descriptions with $\Lambda \neq 0$ appears in the paper he wrote with Einstein in 1932 [87]—otherwise, his articles always accounted for this generality; although, in conclusion to the comprehensive report he wrote, containing his first reaction after learning of Lemaître's article [63] from Eddington, he had clearly not come to the whole-hearted acceptance that Eddington had [72]:

> The constant Λ, which is a measure of the inherent expanding force of the universe,

> is still very mysterious, and it is difficult to see what its real meaning is. It might even be thought to be one constant too many, unless we may hope that it will ultimately be found to be in some way connected with Planck's constant h.

And, in [97], Chandrasekhar wrote that he

> once asked Lemaître, sometime during the late fifties, what, in his judgement, was the most important change wrought by the general theory of relativity in our basic physical concepts. His answer, without a moment's hesitation, was 'the introduction of the cosmical constant!'

Indeed, Lemaître appears always to have been influenced by his time spent at Cambridge, and in 1949 he described 'the cosmical constant as a second constant demanded by the logical structure of [general relativity] theory' [111].

Lemaître's adherence to the cosmological constant may in the end have come from an astronomical standpoint, as he believed throughout the remainder of his career that positive Λ could account for the observed clustering of galaxies in the Universe [76, 77, 78, 111, 112, 113], which remains a little-known—but likely somewhat valid—interpretation. Lemaître first described his theory in 1933 [76], in an analysis which determined the gravitational boundary of a massive cluster. This theory, which he later credited to de Sitter, in [111], who had dedicated a section of [88] to an astronomical analysis of the 'Balance of gravitation and expansive force', seems to have been commonly conceived due to the observed blueshifts of nearby galaxies; e.g., as Eddington described in the above quotation.

Essentially, it was well-known that some nearby galaxies have blueshifted spectra, and therefore from our perspective they do not appear to take part in the greater cosmic expansion, but are gravitationally bound to the Local Group. Because observational evidence exists that such structures dominate the Universe—i.e., many galaxies appear to be members of clusters containing various numbers of galaxies—de Sitter and Lemaître proposed that a similar situation exists throughout the Universe, as in our Local Group, in which there is a boundary at the edge of each cluster where the 'Newtonian' attraction is balanced by the cosmic repulsion due to Λ. In an idealised situation, in which the mass of the cluster is spherically symmetric and non-rotating, the radius of this boundary is simple to calculate, and Lemaître used what is now referred to as the Lemaître-Tolman [114] metric and found that the equilibrium radius of a spherical mass is

$$r_e{}^3 = \frac{3GM}{\Lambda c^2}, \tag{1.13}$$

or, adopting $\Lambda = 1.27 \times 10^{-52}$ m^{-2}, from [115], along with modern values of the other constants,

$$\begin{aligned} M &= 5.70 \times 10^{-26} \text{ kg m}^{-3} \cdot r_e{}^3, \\ &= \left(\frac{r_e}{106 \text{ pc}}\right)^3 \text{ M}_\odot, \end{aligned} \tag{1.14}$$

where 1 $\text{M}_\odot$ is the Sun's mass.[4]

[4]Lemaître adopted $\Lambda = 10^{-50}$ m^{-2}, and thus calculated a denominator of (80 ly)^3, or (25 pc)^3 [76].

In 1954, Felix Pirani proposed a similar theory [116]: he found, equivalently, that the radius r_e in Eq. (1.13) is that of the largest circular orbit about a SdS mass which a test-particle can maintain, and proposed that outside the corresponding sphere—i.e., Eddington's 'sphere of rather more than a million light-years radius round our galaxy'—an orbiting object in any galaxy cluster would necessarily spiral outwards due to the cosmic repulsion. More detailed investigations of the SdS geometry, which include this result, are given without reference to astrophysical implications in [117, 118].

The Local Group's equilibrium radius has recently been determined to be $r_{e_{LG}} = (0.96 \pm 0.03)$ Mpc [119]. Using this value in Lemaître's mass equation, Eq. (1.14), we find $M_{LG} = (7.4 \pm 0.7)\ 10^{11}\ \mathrm{M}_\odot$. The mass determined from this value of $r_{e_{LG}}$ in [119], from an analysis which attempts to describe the boundary within the Einstein interpretation of cosmic expansion, according to which the Hubble flow would tear galaxies away from the perimeters of massive clusters—i.e. according to an imbedding in the overall FLRW Universe, with its best determined parameters—is $M_{LG} = (1.9 \pm 0.2)\ 10^{12}\ \mathrm{M}_\odot$. Due to the vastly different methods from which these two results were obtained, and to the non-trivial assumptions involved in each analysis, this factor of $\lesssim 3$ discrepancy is surprisingly small.

When Lemaître further developed this theory, he noted a duality admitted by Einstein's equation: that Λ can equivalently be thought of as a vacuum energy with negative pressure, with equation-of-state, $p_\Lambda = -\rho_\Lambda$ [78]. This interpretation—which was essentially anticipated by de Sitter [57], who wrote,

> In Einstein's theory of general relativity there is no difference between gravitation and inertia. The combined effect of the two is described by the fundamental tensor $g_{\mu\nu}$, and how much of it is to be called inertia and how much gravitation is entirely arbitrary. We might abolish one of the two words, and call the whole by one name only.

—helped to describe the theory Lemaître would always adhere to, in which systems initially in equilibrium might be perturbed in such a manner that some regions would collapse, becoming separated ever further from each other by intermediate expanding space, driven by an effective negative pressure. A further development, along similar lines of thought, will be discussed in § 4.2.

Historically, the most noted significance of Lemaître's interpretation has been that it provided a link in general relativity theory—especially in the cosmological problem—to a developing concept from quantum theory:—that the vacuum may have non-zero energy; see, e.g., [120, 99] for descriptions of the early history. In the simplest scenario of this theory, in which the vacuum energy would not evolve in time according to some quintessence field, its density, being independent of spatial expansion or contraction, would remain constant; so the vacuum would have an equation-of-state equivalent to that of Λ. Thus, Lemaître had provided the analytical prediction, that, if the vacuum energy were non-zero, it might likely act as a cosmological constant.

At the very heart of Lemaître's interpretation, we have therefore found an answer to de Sitter's rhetoric: although gravity and inertia are formally equivalent in general relativity theory, there remains a basic difference between equivalence and equality, between relative perception and reality, which must always be considered if we care to understand the essential truth of any system described by the mathematical equations.

After these attempts to explain phenomena by positive Λ, few others were made in the remainder of the twentieth century. As well, the concept of vacuum energy would take an equally long time to re-emerge as a central problem of modern physics. In astronomy, Λ would thus slowly fade into obscurity.

The most notable resurrection of the cosmological constant prior to its observation was in the Steady State theory, proposed in 1948 by Bondi and Thomas Gold [121] and by Hoyle [122], which made use of Λ in the same vein as in the Eddington interpretation, and was in fact closely related to Weyl's theory. The Steady State theory, like Weyl's cosmology, would describe the Universe as a slice of de Sitter spacetime—i.e. according to Weyl's postulate—with infinite age, in which the cosmic expansion would be determined solely by Λ. However, rather than having a defined beginning at $t = -\infty$ when all physical distances converge to zero, and exponentially expanding until our Galaxy would appear alone in the Universe after the signals from all others would be redshifted beyond detectability, as Weyl's model required, the Steady State theory operated from an ammended, 'Perfect Cosmological Principle', in which the Universe would appear as it does to all observers in the Universe *for all time*. In order to account for the observed galaxy distribution and expansion, the theory required a steady creation of matter from the vacuum, which would collapse to form galaxies in intergalactic space. It would thus avoid the aesthetical issue that in the infinite existence of the Universe there would be only a short time during which life could exist and the Universe would appear as it does, with the multitude of galaxies we observe, which is a noted downside to most other cosmological theories, especially those of Weyl and Eddington.

This theory was however refuted when the CMB was discovered in 1965 by Arno Penzias and Robert Wilson [123, 124], which confirmed a prediction of the theory on the origin of elemental abundances developed by Gamow, Ralph Alpher, and Robert Herman in the late 1940s [125, 126, 127, 128, 129]. This theory would account for the abundances of the light elements in the Universe—most notably, the roughly 25% abundance of helium—as well as predict the existence of a cosmic background radiation (CBR) produced in the initial hot state when the Universe had sufficiently cooled so that radiation could decouple from matter, allowing electrons to recombine (for the first time) with nuclei, which would subsequently have cooled through the expansion of space to 'about 5° K' [128]. The Steady State theory had trouble accounting for the helium abundance and the CMB, and was therefore abandoned for the Big Bang theory, which naturally accounted for both of these observations.

Shortly thereafter—in 1967,—a cosmological theory involving Λ was again proposed; this time, in order to account for the apparent 'preponderance' of quasars at redshift $z \sim 1.95$ [130, 131, 132]. This theory—call it PS$_3$K for its authors—employed Lemaître's cosmological model, with positive curvature and a cosmological constant, $\Lambda = \Lambda_E(1+\varepsilon)$, $0 < \varepsilon \ll 1$, which described a period of stagnation during which the closed universe would slowly approach the Einstein radius, remaining there (roughly) for an arbitrary length of time—depending on how small one chose ε—before continuing its expansion at an accelerating rate. During this stagnation, light previously emitted by quasars would not be further redshifted, but, as it travelled through quasi-static space, would develop an absorption spectrum after passing through galaxies and interstellar dust. Thus, the theory hypothesised that the stagnation period occurred at a redshift of $z = 1.95$, and sought substantiation from an observation by Geoffrey and Margaret Burbidge [133, 134], that some quasars appear to have emission redshifts which exceed their absorption redshifts of $z \sim 1.95$. However, the PS$_3$K theory never gained wide acceptance, and was completely rejected as more data and a theoretical understanding of the evolution of quasars were brought forth in the early 1970s. But it is historically significant because of its link to the first second-order calculation of the quantum vacuum energy, published later that year, by Yakov Zel'dovich [135, 136, 137].

Zel'dovich [136, 137] used arguments similar to some of those we have already considered, in arguing against the Einstein interpretation of Λ; i.e., he argued that there is no theoretical justification for assuming $\Lambda \equiv 0$;—that appealing to aesthetics and logical economy is a weak form of argument which could never stand up against objective data. He then pointed out, that observations of the cosmos, being always subject to the limits of experimental uncertainty, can never prove Λ is absolutely zero, but could only ever arrive at this result gradually. Here, he made his famous statement regarding the cosmological constant: 'The genie has been let out of the bottle, and it is no longer easy to force it back in.' Historically, this statement would hold true following Zel'dovich's work—for although the PS_3K theory would die, it had done its part in motivating Zel'dovich's calculation, which would eventually find itself as one of the pioneering works on one of the most important problems in modern physics—the cosmological constant problem; see, e.g., [120, 99] for reviews of the problem from this decade.

Interest in the vacuum energy and its possible connection to cosmology grew steadily during the 1970s, following the calculations by Zel'dovich. In particular, symmetry breaking after the hot Big Bang became a popular topic. A review of the early research in this field up to 1975 is provided in *Relativistic Astrophysics, Vol. 2*, by Zel'dovich and Igor Novikov [138]. It describes theories which had emerged suggesting that the vacuum energy might be very large shortly after the Big Bang, due to broken symmetries, in which a link to the cosmological constant had been noted.

Then, while working on the problem of a grand unified field theory in 1980, Alan Guth noticed that some scalar fields might get caught in a local minimum and drive an exponential inflation, from a vacuum energy that would act like a cosmological constant [103] (which references [139]). The inflation scenario, which Guth proposed the following January [140], had many problems, as he noted, and has been revised over the years; a recent description is given in [103]. The importance of inflation theory has however remained unchanged: it proposed a solution to the horizon and flatness problems, which had been relatively unknown, but have since gained notoriety as significant problems of cosmology.

Both problems were discussed in detail by Guth in [140]. Briefly, the horizon problem is, that in the FLRW models alone, in which the cosmological spacetime geometry would be determined only by the distribution of matter, there is no way to account for the observed isotropy in the present Universe. This is because our particle horizon, or spherical region of influence, from which we currently receive light emitted at some time in the distant past, has increased linearly with time, while the FLRW scale factor has gone more like $t^{1/2}$ or $t^{2/3}$, so long as Λ was negligible. Thus, if the expansion of space is to be determined locally by the distribution of matter, as the FLRW models assume, regions we now observe would not have been causally connected in the past, and the present distribution should not be as smoothed out as it appears. The flatness problem is that the Universe appears to have approximately zero curvature. However, according to Friedman's equations, for the Universe to be as flat as it appears today, it would've had to have been unimaginably near to being perfectly flat at very early times, as even very small deviations from flatness would lead to a present Universe which is nowhere near flat. Of course, the flatness problem vanishes if the Universe is exactly flat; but as Guth argued, 'the universe is certainly not described *exactly* by the Robertson-Walker metric. Thus it is difficult to imagine any physical principle which would require a parameter of that metric to be exactly equal to zero' [140].

Inflation would solve the horizon problem if a rapid expansion would have increased the size of our prior causally connected region faster than the speed of light, so that it is still much larger than our present particle horizon. This rapid expansion of space is also supposed to have flattened the curvature, like being on the surface of a sphere of increasing radius; but, as Roger Penrose has pointed out, if the matter distribution were, e.g., fractal, inflation would have had no such effect [104]. A very readable critique of the current status of the inflation scenario was recently given by Paul Steinhardt [141].

Aside from the immediate significance of inflation theory in potentially solving both of these problems, it also had a hidden significance in that it was the first prediction, since the combination of the PS_3K theory and Zel'dovich's work, of a solution to a problem in cosmology that would be given by a theoretical calculation of vacuum energy. However, whereas the PS_3K theory was only short lived, the inflation theory has endured in cosmology, and has thus been a long-standing connection between cosmology and particle physics.

Throughout the 1980s, further analytical work on the vacuum energy from symmetry breakings and experimental verification of the Standard Model of particle physics—e.g. the discovery of W and Z gauge bosons in 1983, which confirmed the electroweak interaction—led to a solidification of the theory, and evidenced a significant discrepancy between all theoretical predictions and the observational limit on Λ from cosmology. In 1989, Steven Weinberg wrote a review [98], outlining the problem in which the limit on Λ was thought to be different from the vacuum energy predictions by anywhere from forty to 120 orders of magnitude, as well as considering its various potential solutions—none of which would prove very promising. And aside from the confirmation of a small cosmological constant in 1998 [49, 50], which is roughly equal to the upper bound that had been known from the observed cosmic expansion, little has changed in the general picture, and no well-motivated theory has emerged with the power to properly explain this discrepancy. In 1983, Novikov asked the question, 'Is the vacuum gravitating?' [142]. Due to the disparity which has only become more sure in the intervening years, it is now more common to hear the phrase, 'The vacuum is gravitating extremely weakly.'

A noteworthy obstacle of the problem, is that, according to the duality of Lemaître's interpretation of Λ, the observed dark energy can only be known as the addition of a pure Λ and vacuum energy. For consider the vacuum Einstein equation with non-zero Λ and with some amount of Λ-equivalent vacuum energy:

$$G_{\mu\nu} + g_{\mu\nu}\Lambda = \kappa T_{\mu\nu,\mathrm{vac}}, \tag{1.15}$$

where $\kappa T_{\mu\nu,\mathrm{vac}} \equiv -g_{\mu\nu}\Lambda_{\mathrm{vac}}$ explicitly defines the vacuum energy as a constant source. Note, that the vacuum energy is a perfect fluid, with $T^{\mu\nu}_{\mathrm{vac}} = (\rho_{\mathrm{vac}} + p_{\mathrm{vac}})u^\mu u^\nu + p_{\mathrm{vac}} g^{\mu\nu}$ and equation of state, $p_{\mathrm{vac}} = -\rho_{\mathrm{vac}} \equiv -\Lambda_{\mathrm{vac}}/\kappa$. Now, we are free to move Λ to the right-hand side of Eq. (1.15) and combine constants, defining an effective cosmological dark energy term, $\kappa T_{\mu\nu,\mathrm{DE}} \equiv -g_{\mu\nu}\Lambda_{\mathrm{vac}} - g_{\mu\nu}\Lambda$, with Einstein equation

$$G_{\mu\nu} = \kappa T_{\mu\nu,\mathrm{DE}}. \tag{1.16}$$

Therefore, regardless of all possible complexities of a potentially non-zero vacuum energy, Λ may always be absorbed into the stress energy tensor as a source. Then, if the vacuum energy is a pure constant, the observed expansion rate and our knowledge of quantum theory tell us the term $|\kappa\rho_{\text{vac}} + \Lambda|$ must cancel to at least forty decimal places.

However, according to quintessence theories, this does not have to be so; for, once we have made this identification, we are free to ask whether the dark energy density really is a constant, or whether it varies slowly at sufficiently late times, so that, in the observable past, it has resembled a cosmological constant, as measurements constrain. Then, even if Λ has a non-zero value, it could act as a potential minimum of the quintessence field. In [98], Weinberg notes that in order for the vacuum energy in the electroweak and earlier transitions to produce a small effective cosmological constant at late times, it is necessary to choose the potential so that earlier, before the phase transitions, the dark energy would have had a very large value, in connection with inflation theory. He has remained a strong advocate of this theory, and, in [103], presents an extensive analysis of quintessence models, which he feels is the surest approach to solving the cosmological constant problem.

Let us now turn to the SN Ia observations, which provide the most direct evidence of the nature of cosmic expansion: SNe Ia are explosions of white dwarf stars in binary systems following sufficient accretion of mass from their companions to reach the limiting mass that can be supported by degenerate electron pressure, known as the Chandrasekhar limit [143, 144], so that their constituent carbon would necessarily ignite explosively. Thus, the luminosity curves of these explosions are always the same, aside from measurable perturbances, and they act as 'standard candles'; i.e., when their apparent brightnesses have been measured, their distances can be calculated quite accurately. By observing numerous SNe Ia at various redshifts, astronomers have therefore been able to fairly accurately trace the changing rate of cosmic expansion, and the result, which has been known for more than a decade now [49, 50], is that the cosmic expansion is accelerating.

From Eqs. (1.9) – (1.10), we have the acceleration of expansion in the FLRW universes,

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{\kappa}{2}\left(p + \frac{\rho}{3}\right). \tag{1.17}$$

Therefore, in any FLRW model, a measurement of positive acceleration requires positive Λ, in order to eventually overcome the decreasing contribution to the expansion rate from matter—for which $\rho,\ p \geq 0$—at late times. In 2007, the ESSENCE ('Equation of State: SupErNovae trace Cosmic Expansion') SN Ia data were found to place significant statistical constraint on the restriction of exotic dark energy models to the so-called flat ΛCDM model, which describes a flat FLRW universe dominated by a pure cosmological constant and cold dark matter [145, 146]. The constraint on dark energy was strengthened in subsequent years, with the publications of results evaluated using the Union [147] and Constitution [148] SN Ia data sets. More recently, a claim was presented in the publications of cosmological results from the first year Sloan Digital Sky Survey-II (SDSS-II) SN data release [149, 150], that these data may favour more exotic dark energy; however, there is evidence that this result is artificial [151].

Measurements of anisotropies in the CMB and baryon acoustic oscillations (BAO) together provide constraints on cosmological parameters which are independent from those found by analysis of SN Ia data, and which have also always been consistent with the flat

ΛCDM model; and, as with the SN Ia data, the most recent results from the Wilkinson Microwave Anisotropy Probe (WMAP) 7-year data [152, 153] and the spectroscopic SDSS Data Release 7 galaxy sample [154] have significantly constrained those parameters.

Thus, the current status of the cosmological constant problem is that quantum theory predicts a vacuum energy which is much larger than the observed value of $\Lambda_{\rm DE}$, and there is yet no convincing evidence that a decay to that value should have occurred. Furthermore, the value of $\Lambda_{\rm DE}$ measured from SN Ia light curves, as well as from CMB and BAO measurements, has the same order of magnitude as the indirect constraint which had previously been placed on it through expansion dynamics; viz., $\Lambda_{\rm DE} \approx 10^{-52}$ m^{-2}. This constitutes half of the 'coincidence problem':—that the matter and vacuum densities are roughly equal at present. For if $\Lambda_{\rm DE}$ were truly unrelated to the primordial cosmic expansion, it might instead have been that $H_0 \gg \left|\sqrt{\Lambda_{\rm DE}}\right|$, rather than the two terms being nearly equal; then, even with the precision measurements from SNe Ia and the CMB, $\Lambda_{\rm DE}$ would have gone undetected. On the other hand, $\Lambda_{\rm DE}$ might have been very large, so that the Universe would have quickly become de Sitter's vacuum, with matter density negligible in comparison, and the Universe would now appear much different to us; e.g., if galaxies *could* have formed, they may all have vanished from sight long ago, and we would never have detected the Universal expansion. Such possibilities motivate anthropic arguments, as, e.g., Weinberg has considered [98].

Altogether, these results suggest that the measured value of $\Lambda_{\rm DE}$ is not likely related to the predicted vacuum energy from quantum field theory, but that our Universe has a true cosmological constant—which could rather be essentially linked to its apparent expansion. If this were true, the implication could be that we must rectify quantum field theory with vanishing vacuum energy, or else our physical cosmology would have to require that the large vacuum energy does not contribute to the cosmic expansion.

Now, the principal issue with any attempt to reconcile Eddington's interpretation of expansion with theory, is that it is inconsistent with FLRW big bang cosmology, according to Eq. (1.17). As we shall eventually see, however, the RW line-element corresponds only the most trivial first guess that may be made in order to formulate a dynamical model to account for our cosmological observations. And this first guess, which should anyhow be useful for modelling those *observations* from anywhere in the Universe, according to our own observations and the cosmological principle, has since been investigated extensively, with the uncomfortable result, that when we further interpret the observational model as a representation of the Real physical state of the Universe, the theory is plagued with many significant problems to which we have been unable to find truly sufficient solutions.

Most notable among these, is the frequently overlooked problem that was known to Eddington:—that, in the big bang scenario, FLRW cosmology cannot *explain* the expansion we use the model to empirically *describe*; for only Weyl's 'de Sitter cosmology', completely devoid of matter, is capable of explaining the expansion from the moment of a big bang through a positive cosmological constant, while all other FLRW big bang universes require an indescribable singular push in order to exist. Thus, these models can describe only the effects of extrinsic sources on a primordial expansion rate, and provide no realistic clues as to its origin.

The Nobel Prize-winning discovery late last century, that the cosmic expansion is accelerating, was therefore a nuisance for standard cosmology because it adds a complication to the logical simplicity of the theory; viz., that the measured positive $\Lambda_{\rm DE}$ has no fundamental role in the theory, being unnecessary for the existence of a FLRW big bang universe, and so begs the question, 'Why should it exist at all?' Λ_{DE} has therefore been popularly thought of as vacuum energy, which could have had a role in the inflation scenario if the quintessence theory is correct. Otherwise, only at late times would pure Λ become relevant for describing cosmic expansion.

In essence, in the standard theory, the observation of cosmic acceleration is an annoyance, and the measured value of Λ_{DE} is a problem. And the logical simplicity that Einstein had sought while supporting the Einstein-de Sitter universe was in fact only superficial, as it imposes on cosmology an impossible complication, in which the entire history of the Universe is ultimately to be the result of an inexplicable singularity, at which point (we are then forced to conclude), the theory itself must fail to describe things correctly; thus, Eddington's interpretation is logically much simpler because, at the very least, it does not need to presuppose that the mathematical theory needs to break down at the single event that is also supposed to essentially give rise to our extraordinarily particular reality (for more discussion on the extraordinary specialness of the particular big bang that this theory should require, see [104, 105]).

It is no wonder that Eddington was so incensed with the position Einstein had finally taken regarding the observed expansion, which came only after Hubble's confirmation, after he had remained silent on the issue during the prior decade's theoretical investigation by so many others: the feeling surely would have been similar for Gamow, Alpher, Herman, and Lemaître, if, in 1965, Robert Dicke and his colleagues had, instead of [123], written that Penzias and Wilson's discovery confirmed a prediction of the Steady State theory, in the particular case in which God continually fills the Universe with a steady amount of one-quarter helium, three-quarters hydrogen, and three-degrees of microwave radiation; for the *explanation* of expansion, according to a prior physical cause, was similarly unjustified in the Einstein-de Sitter model, which not only completely neglected the manner in which the observed expansion had been *anticipated* through physical theory, but even neglected any mention of the possibility.

Einstein seems to have had little interest in understanding prior physical causes, as a means of developing a general relativistic cosmological theory that would reasonably account for the observed phenomena: instead, he seems to have been primarily concerned with the construction of an empirical model that would fit the observations in a manner that would remain consistent with his previous reasoning—i.e., he really didn't give much further thought to the problem of expansion. His evident lack of interest in finding a prior explanation to this problem is indeed a concern; and the best explanation that I can see for this apparent apathy, is that he had already led himself towards more abstract ways of thinking, due to the idealist-determinist interpretation of the meaning of relativity that he always considered to be the most objective one, which led him to accept, sometime long before the 1930s—e.g., as already evidenced in the postcard he sent Weyl in May, 1923,—that the general theory must be essentially flawed.

The abstraction of thought to which I am referring is evident, e.g., in his autobiography, where he wrote that he spent seven difficult years (1908-1915) freeing himself 'from the idea that co-ordinates must have an immediate metrical meaning' [40]. However, in contrast to the way of conceiving things he thus came to, the coordinate systems used to describe physical reality *do* have a clear metrical meaning in the dynamical theory of relativity proposed in

Chapter 3, which describes the relativistic spacetime of empirical observation as an ideal graph of events which occur in an objectively well-defined three-dimensional universe that physically evolves in an objectively well-defined manner.

But so our investigation must first turn to the problem of understanding why Einstein had denied this possibility, and accepted instead an abstract interpretation of the relativistic description of events in physical reality that ultimately requires four-dimensional block determinism; and we shall see that he was in fact entirely consistent in that interpretation—and, therefore, that the common application of relativity theory, which has always demanded the same interpretation of the mathematics that *does* formally require spacetime to be a singular block eternity, has not been so consistent. Accordingly, then, it shall become clear that even though relativity theory is commonly used to describe dynamic processes, the inherent inconsistencies in such applications have obstructed an accurate understanding of the true dynamical description that is afforded by the theory, and many paradoxes have subsequently been inferred which we shall immediately see resolved when the formally consistent dynamical interpretation of the mathematical theory is assumed.

So, in order to completely justify that interpretation as the most objective one, we must thoroughly assess the many aspects of this basic problem of four-dimensionalism, and try to understand the assumptions and presumptions behind the reasonings of those who have played important roles in its philosophical development, with the idea that if we can truly understand what they were thinking and why, we might better understand where they went wrong. For only then, when we have subsequently applied that knowledge in consistently working through the formal requirements for causal dynamism, will we be able to properly address the cosmological problem of relativity theory in a straightforward, logically consistent manner, and work out a solution that is consistent with cosmological observations, the mathematical theory, and the true dynamical nature of reality we want our theory to describe—which, we'll find, objectively accounts for the details demanded by both reason and observation.

Chapter 2
Against Einstein's Relativity: A Doxographical Analysis

> I cannot believe—and I say this with all the emphasis of which I am capable—that there can ever be any good excuse for refusing to face the evidence in favour of something unwelcome. It is not by delusion, however exalted, that mankind can prosper, but only by unswerving courage in the pursuit of truth.
>
> —Bertrand Russell, *Fact and Fiction*

2.1 Introductory Remarks

The primary goal of this investigation, maintained from the outset, has been to find a natural description of reality that is implicitly consistent with our observations—not by trying to rectify any apparently paradoxical or conflicting facts within an assumed framework, but by attempting to identify and analyse together all the basic facts that we can know objectively, and using them as indicators of what would therefore seem, prior to forming any prejudice, even according to any logical reasoning, to be the most natural interpretation; i.e., not to allow only for naïvely objective deduction of only certain specific facts, neglecting that the resulting theory might ultimately be at odds with certain others, and then opting to ignore such problems, but also to base any subsequent deduction on some common sense and basic observational facts, which are relevant to our description of reality.

The idea is, that in specific instances regarding our observations of reality, multiple interpretations can be made logically,—some of which do not however fit naturally into the greater picture, as inconsistencies are later discovered. Logically, therefore, in the grand scheme, such interpretations are therefore found not to fit into a natural description of reality. It might be thought, that if the whole picture could be considered objectively, such paradoxical interpretations would never enter into our theories, as they would be immediately removed by such logic. We are thus led to consider Laplace's demon [155], who would know, e.g., which interpretation of quantum theory is actually correct, by considering and comprehending all of reality at once. Because it is beyond human capability to do this, we are limited to making logical interpretations of more specific observations.

But this is inherently problematic; for if we would, e.g., regard specific observations always with complete independence, then we should never be able to decipher which, of the many logical interpretations that would be allowed, would actually be the correct one; and even if we lessen the extremeness of this requirement, so that different objective measurements would be considered together in our deductions, we can never consider *everything*; so we cannot determine Truth—if such exists, which we take *in principle* to be so—purely through objective deduction in considering the logical interrelatedness of multiple observed facts.

On the other hand, we have our sensory perception of Nature, which, although a fact of observation, is difficult to rationally inspect; in particular, it is often difficult to distinguish the conclusions that are drawn intuitively from such perceptions—and, indeed, whether those are not perceived at all, but conceived only by the mind—from the basic facts of perception. Regardless, we must lay some trust in our own basic sensory perceptions, and, if we have seen that they are common ones, suppose that they may result from some property of Nature we are thus observing, and allow them to guide our interpretations of reality while keeping the limitations of human perception and understanding in mind; for the facts contributing to common sense seem to be critically important, and may be unobservable otherwise.

Therefore, we say that when a scientist makes some observation, which could be logically interpreted many different ways, they are not being properly objective and truthful if they would say, 'The truth could as well be any of these things, which I therefore shan't commit to any further,' if they know that certain of these interpretations are only allowed because the observation is being considered singularly, or even in conjunction with some others, but in denial of some aspect of common sense, merely brushing it aside; for if they would be truly objective, they would consider together all relevant observations, including basic sensory perception, and make their interpretation based on whichever thus seems most logical; for, along with the many other principles to which one must continually remain consistent, in order to be truly objective one must also be true to oneself.

It must be noted that any such examination of beliefs and biases, which aim is to discover their Natural foundations and include them in physical theory, is very different from yielding to common prejudice in theoretical development—whether that prejudice initially came as the result of critical thinking or otherwise; for only a theory that is consistent with *all* empirical data is worthy of any amount of such prejudice, and then only so long as no subsequent inconsistencies are discerned, and only to the point that its basic principles can be accepted as the True facts of Nature.

The problem on which the next two chapters are centred, is that of the True metaphysical nature of Time, and how that is described in relativistic cosmology. For the true nature of Time has long been a topic of study in natural philosophy, and has by no means been resolved by relativity theory—the standard interpretation of which—if any one standard can really be said to exist—seems to lead to more paradoxes and inconsistencies than resolutions; therefore, in order to arrive at a general relativistic theory that could be used to describe the large-scale structure and evolution of our Universe, an appropriate view of the nature of time, and how that is described by the standard spacetime formalism of general relativity theory, must first be developed.

When this has been accomplished, the theory will eventually resolve as one consistent with the mathematical formalism of relativity theory, but in which the many well-known inconsistencies and problems, the apparent paradoxes which have commonly led to *ad hoc* conjectures, etc., identically vanish. It is, by the way, the only theory of time, by which the dynamical cosmology described in Chapter 4 makes any sense; but also the one by which the structure, evolution, and *origin* of the Universe has potential to be truly explained *a priori*, rather than given, as it is with the standard cosmology, as a description *a posteriori*, which, although more general in application, is no more than that, and, in fact, which is the result of an interpretation of relativity theory requiring that the Universe does not actually *evolve* at all, but is laid out, once and for all, as a singular four-dimensional continuum of events, which does not admit any *real* dynamics.

The contrary picture I will therefore describe—which must be described, because it is the only one through which any meaningful *understanding* can come of the foundational principles of relativistic cosmology—which, when stated below, will correctly be understood to be only superficially and trivially satisfied by the standard theory,—is not altogether new. In fact, it is closely related to the cosmological theory that Weyl developed in the early-1920s, over which he did battle with Einstein [22], and also to a theory that was eventually reasoned out more than a millennium earlier, by St. Augustine—which is really only extended here, by requiring a *physical* fourth dimension of reality in that theory, and adding to it the formal mathematical framework of relativity theory.

Although an argument for that theory must ultimately be given, the greater goal of this analysis is that it should provide a clear *explanation* of this one theory. But in order to give a clear account of the meaning of this theory, and thus explain the general idea of what it is, which ultimately requires knowledge of what it is not, the discussion cannot be a complete logical deduction of all possibilities, which would ultimately obscure the straight and narrow path that is necessary in order to meet that specific purpose. Therefore, although an objective deduction of all the facts that are known, or may be immediately inferred, as well as the various philosophical theories that have been proposed throughout history, must be possible, the argument that is so vital to the explanation, is presented by means of whatever method of logical inquiry seems to be the most direct means to that end, at each point, so that we are able to maintain logical consistency while addressing only those points which seem to be the most pertinent to the problem.

In order to properly present such an argument, it is important to start at the beginning; for without a proper account of the rudiments of our modern theory, only a shallow explanation is possible; as any one wishing to fully understand a theory must know the reasoning behind it. Furthermore, it is at once obvious that in analysing the very roots of discovery, objective hindsight and modern intelligence should serve to dispel any errors that might have entered into a theory through contemporary prejudice, or to clearly see at which points subtle choices were actually made as such, when only one clear path may have been originally noticed—or, in general, to navigate the way along a path which was not originally taken, but which now seems to be more fruitful, as the most conducive to the discovery and formation of a better understanding of Nature.

For in these regards we find ourselves in common sentiment with one of the greatest physicists in history, who wrote that 'It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in the nascent state' [156]. For it is not from the epitomes of the original arguments, but from the prior reasonings through which those conclusions were drawn, that we stand to gain the most insightful perspective from which to improve on our understanding. In western philosophy, it appears that the first meaningful speculations about the Nature of Time came from opposing viewpoints, described by the Eleatics, who followed Parmenides, and by Heraclitus, who both flourished just prior to the turn of the fifth century B.C., and whose works now survive only in fragmentary quotations and secondary accounts, as with all of the pre-Socratic philosophers.

Before addressing the relevant aspects of these two theories, which we'll do in §§ 2.2 – 2.3, I think it should help to clarify the greater scope of this chapter and the next—and further, to see precisely how important to modern theory the works of Heraclitus and the Eleatics really are—to contrast it with a statement to which it is intimately linked, made by Sextus Empiricus, whose collective works, written in the late second century A.D., are an important source of pre-Socratic philosophy [157] (III.65):

> The most fundamental positions on motion have, I think, been three in number. Common sense and some of the philosophers suppose that there is such a thing as motion; Parmenides and Melissus and some others think that there is not; and the Sceptics have said that there no more is than is not—so far as appearances go there seems to be motion, so far as philosophical argument goes it is unreal.

Sextus then sets out to show that the stance of the Sceptics is the right one.

In contrast, I shall maintain throughout, that the Sceptics' viewpoint is perfectly hypocritical. I say 'maintain', because I will not attack this view directly, but will carefully analyse the other two theories, as the only two formal possibilities, and eventually discuss their relevance to general relativity theory. And I say 'hypocritical' because the Sceptical stance, as it is here described, is *essentially* Eleatic, but it pretends not to be the same as Eleatism, based on a concession that motion must be apparent, even though this was also recognised by the Eleatics. In principle, one must admit that a philosophical theory cannot *be* one way and *appear* another way, without still *being* what it essentially *is*.

So it is, that Sextus has really only recognised two different essential positions. If any meaningful distinction could be said to exist between the Eleatic and Sceptic views, as he has tried to say, it must be only, that on this point Scepticism regresses from the sophistication of its predecessor, and forgets what it is *in principle*, which naïveté inevitably gives rise to the acceptance of apparent paradoxes as genuine, yet abstract, truths—e.g., 'that there no more is than is not' motion.

This can be related to more recent philosophic and scientific theories, beginning with Immanuel Kant's transcendental, or critical idealism: viz., as Kant correctly realised, nothing of reality can be observed, except its *ideal*[1] aspects; so he refused to consider any other possible aspect of reality, other than that which may be subjected to science. Kant did not deny the existence of a metaphysical reality which would give rise to the observed *ideality*, but, since no scientific description can be ultimately dissociated from an underlying metaphysical world-view, in denying the importance of describing and understanding the basic metaphysical noumena by pure reason, and contending that truth is only in experience, his theory becomes effectively equivalent to the philosophies of later idealists, such as Francis Herbert

[1]This is meant in the adjectival sense of *idea*, as in the French *idéel*, or German *ideell*.

Bradley, John Ellis McTaggart, and Einstein, who ultimately committed to the metaphysical view that the ideality of experience is a four-dimensional block Reality.

For, in Einstein's case, he did recognise the fact, that according to the interpretation of the relativity of simultaneity in which the simultaneous realities of two relatively moving observers are supposed to exist concurrently, as formally distinct sets of nevertheless real events, not only do those particular two realities coexist, but so too must the entire spacetime continuum of events, as a strictly determined block, which is all that may be investigated by science. As Louis de Broglie once put it [40],

> In space-time, everything which for each of us constitutes the past, the present, and the future is given in block, and the entire collection of events, successive for us, which form the existence of a material particle is represented by a line, the world-line of the particle. Moreover, this new conception defers to the principle of causality and in no way prejudices the determinism of phenomena. Each observer, as his time passes, discovers, so to speak, new slices of space-time which appear to him as successive aspects of the material world, though in reality the ensemble of events constituting space-time exist prior to his knowledge of them. Although overturning a large number of the notions held by classical physics, the special theory of relativity may in one sense be considered as the crown or culmination of that physics, for it maintains the possibility of each observer localizing and describing all the phenomena in the schema of space and time, as well as maintaining the rigorous determinism of those phenomena, from which it follows that the aggregate of past, present, and future phenomena are in some sense given *a priori*. ...

As discussed in § 2.7, this was well-known to Einstein, and a formal argument has been given in special relativity theory by Wim Rietdijk [7] and Hilary Putnam [8]. For a more general discussion, one may refer to Kurt Gödel's paper in [40] as well, which had previously examined the problem in relation to more general cosmological models. Some more recent investigations can be found, e.g., in [158].

Of course, not everyone has understood that, aside from the 'arrow of time' which provides its causal structure, the standard spacetime formalism should describe nothing more than points in a four-dimensional block. Therefore, by often advancing forms of subjective idealism that mirror the philosophy of George Berkeley, other descriptions of the theory seem to have regressed from the scientific idealism of Kant, formalised by the mathematical theory, which actually requires strict determinism after the standard interpretation of the relativity of simultaneity has been assumed; e.g., this is the case with the descriptions given by Eddington in many of his works,—as in *The Nature of the Physical World* [13], where he provides a paltry description of the meaning of relativity theory involving a subjectively-defined Here-Now and its corresponding light-cone, along with an imprecise neutral wedge, which he called Absolute Elsewhere—which he illustrated with the following hypothetical example:

> Suppose that you are in love with a lady on Neptune and that she returns the sentiment. It will be some consolation for the melancholy separation if you can say to yourself at some—possibly prearranged—moment: 'She is thinking of me now.' Unfortunately a difficulty has arisen because we have had to abolish Now.

> There is no absolute Now, but only the various relative Nows differing according to the reckoning of different observers and covering the whole neutral wedge which at the distance of Neptune is about eight hours thick. She will have to think of you continuously for eight hours on end in order to circumvent the ambiguity of 'Now.'

An example which illustrates the problem with such a subjective idealist scenario is the 'Andromeda paradox', stated by Penrose in *The Emperor's New Mind* [17], according to which a group of Andromedeans might, in one frame, be considering whether to invade the Earth, with the ultimate decision being yet undetermined, while, in some relatively moving frame, they could already be on their way.

But even some descriptions which don't enter into discussion about the paradoxical consequences that come with thinking of spacetime as anything less than a solidly determined block, often do not say explicitly that this must be so. For example, in *The ABC of Relativity* Bertrand Russell came the *closest* to saying this only at the end of his chapter on spacetime, where he wrote [159],

> We may now recapitulate the reasons which have made it necessary to substitute 'space-time' for space and time. The old separation of space and time rested upon the belief that there was no ambiguity in saying that two events in distant places happened at the same time; consequently it was thought that we could describe the topography of the universe at a given instant in purely spatial terms. But now that simultaneity has become relative to a particular observer, this is no longer possible. What is, for one observer, a description of the state of the world at a given instant, is, for another observer, a series of events at various different times, whose relations are not merely spatial, but also temporal. For the same reason, we are concerned with *events*, rather than with *bodies*. In the old theory, it was possible to consider a number of bodies all at the same instant, and since the time was the same for all of them it could be ignored. But now we cannot do that if we are to obtain an objective account of physical occurrences. We must mention the date at which the body is to be considered, and thus we arrive at an 'event,' that is to say, something which happens at a given time. When we know the time and place of an event in one observer's system of reckoning, we can calculate its time and place according to another observer. But we must know the time as well as the place, because we can no longer ask what is its place for the new observer at the 'same' time as for the old observer. There is no such thing as the 'same' time for different observers, unless they are at rest relatively to each other. We need four measurements to fix a position, and four measurements fix the position of an event in space-time, not merely of a body in space. Three measurements are not enough to fix any position. That is the essence of what is meant by the substitution of space-time for space and time.

But then, unless one is prepared to accept very abstract notions which muddle the formal description of Physical Reality, in order to reconcile relativity theory with common sense, the formal requirement of a block spacetime of *a priori* given events must be taken to be at the heart of Russell's recapitulation, even though he appears there only to be beating about the bush.

In fact, it is my opinion that the reason why so many such descriptions of relativistic spacetime have appeared muddled and hard to grasp, by anyone seeing the theory for the first time, is that the describers—whose point-of-departure is always to interpret the relativity of simultaneity as meaning, that two events which appear to any observer to have occurred simultaneously *really* did take place simultaneously in that observer's frame; and equivalently so, for all observers, according to the principle of relativity—do not begin, before moving on to describe the way of 'happenings' in spacetime in more detail, by stating that spacetime, according to the common interpretation of the mathematical formalism, is a strictly determined four-dimensional block, with the particular (somewhat peculiar, even) property that we commonly refer to as the 'arrow of time'.

Now, it is not my intent to show here that this is so, because that has already been done sufficiently well by others—e.g., by those referenced above. In fact, as I *will* eventually point out, in § 2.7, it was in the basic design of the formalism originally developed by Hermann Minkowski in 1908 [6]. What I do intend to show, in Chapter 3, after the more significant elements of this basic problem have been analysed in the current chapter, is that there is a *more* objective interpretation of the mathematics, which is completely at odds with the original interpretation that was made, in effect, by Einstein—which he thought, incorrectly, to be the most objective one possible; for this other possibility offers a clear account of the events of spacetime in an explicitly dynamical way, which can be described through an implicit property of spacetime which is empirically supported.

The dynamical interpretation of relativity theory, which is really only a small advance on the basic idea that is described in [160], will not be a difficult one to accept when it is shown to be entirely consistent with the mathematical theory of relativity, and is the *only* interpretation of the mathematics which admits an *essentially* dynamical description of the unfolding of events, because it is just the interpretation that is usually thought of anyway, even though true dynamism has not been reconciled with the theory in all essential details. Even so, it will only become clear through the basic interpretation that is described in Chapter 3, precisely how it is possible that, as recently stated by George Ellis [28], the 'unchanging' 'Block Universe idea, representing spacetime as a fixed whole, [which] suggests the flow of time is an illusion[, according to which] the entire universe just is, with no special meaning attached to the present time', 'is best replaced by an evolving block universe [view of spacetime] which extends as time evolves, with the potential of the future continually becoming the certainty of the past; [that] spacetime itself evolves, as do the entities within it',—by explaining that even this picture should not be an accurate representation of the *reality* it is used to describe, according to the most objective interpretation of the theory.

However, because the difference is purely ontological, since the mathematical description of physical events upon which it is based remains exactly the same, the argument for real dynamism shall inevitably come down to an appeal to such natural epistemological requirements as Gottfried Leibniz' 'principle of sufficient reason'—viz., that there has to be a reason for everything which is sufficient to explain why it should be so, and not otherwise [161].[2]

[2]In the same paragraph in his second correspondence to Samuel Clarke, Leibniz offered two separate statements of the 'principle of sufficient reason': 'that nothing happens without a reason why it should be so, rather than otherwise', and 'that there ought to be a sufficient reason why things should be so, and not otherwise' [161]; the above statement is supposed to incorporate the meaning of both of these.

And so, in order to provide an adequate argument in favour of the dynamical theory, it is imperative to investigate the essential points that have been made in the historical development of spacetime theory, because, as the above discussion has been meant to indicate, the original interpretation of the theory has had a long history before Minkowski and Einstein, and so has the one I am going to explain here, and there are many subtle points that must be highlighted before the basic justification and reasoning behind either can be completely (*ergo* sufficiently) understood. For only by truly understanding the logical arguments pertaining to either theory, which are best assimilated in their nascent form, where the initial justifications can be properly dissected in order to objectively address their validity, can one free one's mind from the biased thinking that was noted by Einstein as he expressed his opinion in regard to the significance of studying not only the methodology, but also the history and philosophy of science [162]:

> So many people today—and even professional scientists—seem to me like somebody who has seen thousands of trees but has never seen a forest. A knowledge of the historic and philosophical background gives that kind of independence from prejudices of his generation from which most scientists are suffering. This independence created by philosophical insight is—in my opinion—the mark of distinction between a mere artisan or specialist and a real seeker after truth.

As I've said, the theory of physical Reality that I will eventually come to, incorporates the same idea of spacetime that was developed, e.g., in [160], and basically adds to that description an unobservable noumenon; which must therefore be shown, by means that do not directly concern the observation of phenomena, to have clear physical and empirical motivation. Specifically, I will show that it is totally at odds with the very crux of the original formalism of relativistic spacetime theory, in a manner which closely resembles the advance on the Eleatic theory that was made by the Ancient Atomists—and therefore, according to the facts that may be logically deduced, is necessarily required in order to formally reconcile relativity theory with real dynamism.

So, by making use of pertinent insights that many of the most brilliant natural philosopers in history have made, as they previously explored the Nature of Reality, while further clarifying their arguments, or advancing new arguments for or against those, by utilising the advantage of modern intelligence, a clear explanation of the present theory, which probes its basic principles, shall be presented. Then, it will be simple to see that although the theoretical description of physical observables remains exactly the same, many of the formal consequences should differ significantly from the standard conclusions that have been drawn from the original interpretation, when it has been misinterpreted as a dynamical one. As such, many paradoxes that have been previously inferred will be immediately resolved.

This concludes the introductory statement of the goals and methods of this chapter and the next. We now continue our argument by examining the theoretical achievements of Heraclitus of Ephesus, *c.* 500 B.C., beginning with his contribution to epistemology, which is a necessary primer to truly realising his significant contribution to natural philosophy, and is also an essential part of the method that will be used below, as it has been similarly used by many others thoughout the past two and a half millennia.

2.2 On Heraclitus

2.2.1 His Philosophy of Knowledge

As with everything else we know of Heraclitus' philosophy, his epistemology is a topic of some contention. For example, in Sextus Empiricus' *Against the Logicians* [163] (126-134), the brief section on Heraclitus seems to me to begin by stating one thing about Heraclitus' epistemology, but end in showing that it is otherwise. Specifically, he begins his account with [164], '(126) And Heraclitus – since he ... supposed that man is furnished with two organs for gaining knowledge of truth, namely sensation and reason – held ... that of these organs sensation is untrustworthy, and posited reason as the standard of judgment.' From everything that is now known of Heraclitus' philosophy, including the fragments that were given to us by Sextus in this very account, it seems more likely that Heraclitus did not even hold sensation and reason to be on the same level, as paths towards knowledge and understanding, but thought of sensation as the primary form, as our most direct connection to a real underlying Truth, which cannot, however, be correctly understood without clear subsequent rationalisation. As far as I can tell, this is roughly what is described by Sextus below this statement, and it is surely the message contained in the handful of other fragments that have been found elsewhere.

It is therefore a wonder that his next statement is such a clear misconstruction of the meaning of Heraclitus' words, as, according to his own eventual conclusion, it must be regarded as grossly misleading and inconsistent. What he says, viz., is that [164]

> <The claim of> sensation he expressly refutes with the words, 'Poor witnesses for people are eyes and ears, if they possess uncomprehending (literally, 'barbarian,') souls,' which is equivalent to saying, 'To trust in the *non-rational senses* is a mark of uncomprehending souls.'
>
> (127) Reason, on the other hand, he declares to be the judge of truth – not, however *any* sort of reason you might care to mention but that reason which is 'common' and divine. ...

According to what I've already said, however, and as we'll now see more clearly, the meaning of the former statement is not at all equivalent to that of the latter one. Rather, what Sextus should have said was that Heraclitus' statement is equivalent to saying, 'To trust in non-rational *conceptions*—those being directly subsequent to sense-perception, without having been subjected to reason—is a mark of barbarian souls.' Logically, Sextus' statement would have been more closely related to what Heraclitus wrote, if the original statement had been instead, 'Poor witnesses ..., *for* they possess ...'; but this is clearly not at all what Heraclitus thought, as his view was that we *can* reason through to the causes of common sense, in order to discern Truth. The remainder of Sextus' argument would then have been far clearer if the sentence beginning paragraph (127) had been instead, 'Reason, he declares to be the judge of truth—not, however, *any* sort of reason you might care to mention, but reason *about* that which is 'common' and divine,'—which has a very different meaning from that which he did say.

If it is correct to say that this is basically what Heraclitus' theory had been, then we would say that he believed in some common and divine Truth, with which we must interact in some way, in order that we might grasp it in some form, and that we should subsequently subject that which we do grasp to rational inquiry, in order to sort out what it actually, basically, is. If so, it is that which he describes as being commonly grasped, or conceived by the subconscious, that is called common sense, which must indeed occur prior to rational inspection.

In fact, this is the message that must be understood from the remainder of Sextus' account. He goes on to describe what we should now consider to be a mystical scheme of Heraclitus', for how we interact with the 'common' and divine, by breathing it in, whence this comprehending reason is given its divine birth, and eventually concludes [164],

> (131) Heraclitus says, then, that this 'common' and divine *logos* – by participation in which we become rational – is the yardstick of truth. So that which appears <such and such> to all in common is trustworthy (for it is grasped by the 'common' and divine *logos*), but that which strikes only an individual as <such-and-such> is – for the opposite reason – *un*trustworthy.
>
> (132) Thus the aforementioned man begins his work *On Nature*, and in a certain fashion points out <the existence of> the surrounding <nature> with the words:[3] 'And of this account (*logos*) which is the case always men prove to be uncomprehending, both before they hear it and once they have heard it. For although everything comes about in accordance with this account (*logos*), they are like inexperienced men when they experience both the words and the deeds of the sort which I recount by dividing up each thing in accordance with its nature (*phusis*) and saying how it is; but other men do not notice what they do when they are awake, just as they are oblivious of things when asleep.'
>
> (133) For having in these words expressly stated the view that we do and apprehend everything thanks to our participation in the divine *logos*, he goes on a little further, then adds: 'That is why it is necessary to follow that which is <common>. Though the account (*logos*) is common, however, the many live as though they had a private understanding'. This *logos* is nothing else than an explanation (exposition, articulation, *exēgēsis*) of the mode of arrangement of the universe. That is why we speak truly whenever and in so far as we share in the recollection of it (ie, of the *logos*) but are invariably mistaken on matters of private opinion.
>
> (134) So here and in these words he states clearly that the common *logos* is the yardstick <of truth>; the things that appear such and such *in common* are trustworthy, as being judged by the common *logos*, whereas those that appear <such and such> to each person privately are false.

So he ends his account of Heraclitus' doctrine, basically by saying that he had believed in the

[3] I've chosen to use Jonathan Barnes' translation of this fragment, from [165], rather than the one from [164], because it retains the ambiguity of the application of 'always' here, which is commonly agreed upon to have been deliberately constructed, so that a comma could be thought to be placed before or after 'always' in order to mean either 'which is the case always', or 'always men prove to be uncomprehending [of it]'; whereas the latter translator, who agrees with this interpretation, has nevertheless explicitly written 'forever' twice. Furthermore, it seems that 'always' accounts better for the original meaning of the statement, by expressing a generality of ways, than does 'forever', which is strictly temporal.

primacy of common sense, which we must trust to, in order to reason correctly, and only by rationalisation that is faithful to that common *logos*, can we come to a better understanding of Nature's Truth—i.e., expose the True mode of arrangement of the Universe, which he has already assumed to have a prior Real existence.

In summary, this account is understood to say that according to Heraclitus, common sense must be used in order to realise what is True, but its origin *in* Truth cannot be understood unless it is considered with divine reason; i.e., rational, insightful reason, that is always held to be consistent with what is common, as it probes its own very roots. In other words, observations and logical argument must be used in order to find the prior causes of that which everyone knows through common sense. In essence, Heraclitus placed common sense prior to logical inquiry, for it must be the result of something implicit in Nature, but he says that only those who think carefully about it will come to any clear understanding, whereas barbaric misunderstandings result from being careless. In this way, Heraclitus began his account *On Nature* with a direct challenge to all who seek to understand It, that they must always follow this path, for it is the only one through which Truth can be properly exposed.

Many of the other (~ 130) remaining Heraclitean fragments attest to this reconstruction of Sextus' account [164]:[4]

> 17. Many people do *not* 'understand the sorts of thing they encounter'! Nor do they recognise them <even> after they have had experience <of them> – though they themselves think <they recognize them>.
>
> 34. Uncomprehending <even> when they have heard <the truth about things?>, they are like deaf people. The saying 'absent while present' fits them well (literally, 'bears witness to them').
>
> 47. Let us not make random conjectures about the most important matters.
>
> 95, 109. It is better, [says Heraclitus,] to conceal ignorance.
>
> 112. Sound thinking <is> a very great virtue, and <practical> wisdom <consists in our> saying what is true and acting in accordance with <the> real constitution <of things>, <by> paying heed <to it>.
>
> 113. Thinking is common to all. [This is given in conjunction with,]
>
> 114. Those who <would> speak with insight must base themselves firmly on that which is common to all, as city does upon <its> law – and much *more* firmly! For all human laws are nourished by one <law>, the divine <law>. For it holds sway to the extent that it wishes, and suffices for all, and is still left over.
>
> 116. All people have a claim to self-knowledge (literally, 'self-ascertainment') and sound thinking.

And more specific to sense-perception,

> 28a. The most esteemed <of people> 'ascertains' – and holds fast to! – what <merely> seems <to be the case>.

[4] The numbers given here are standards.

46. [He used to say that] thinking is <an instance of the> sacred disease [and that] sight is deceptive.

54. An unapparent connection is stronger (or: better) than one which is obvious.

55. Whatsoever things <are> objects of sight, hearing, <and> experience – these things I hold in higher esteem.

72. They are separated from that with which they are in the most continuous contact.

101a. Eyes are more accurate witnesses than are [the] ears.

123. <a thing's? (the world's?)> real constitution [according to Heraclitus] has a tendency to conceal itself.

This is what we will take Heraclitus' epistemology to have been, which is in total agreement with the stance I have taken in the present analysis, as I have attempted to state in the introduction. Accordingly, we will have no more use for any theory reached by the pure reasoning based on any set of facts, which is not also basically consistent with common sense, than we will have for a common sense-based conclusion that is arrived at without sufficient reason. As such, our main problem will be, as it always has, and may always be,—that we must *only* trust in the completeness of our rationalisations *until* they should show themselves as having been objectively inconsistent, or lacking sufficient reason. However, we find that the surest way to avoid error, as Heraclitus said in so many words, is to use rational insight in order to derive clear expressions of the principles from which our theories are built; for only by clearly recognising what those basic principles are, which must be at once consistent with common sense and all other observations, can we truly understand our theories enough to be capable of objectively assessing their strengths and weeknesses. In other words, his epistemology is based upon a single principle, that all theories come from notions that originate in observation, which bears witness, in some fashion, to an underlying Truth; but in order for any theory to be considered valid (though ultimately refutable, as we can never observe, nor think of, all possibilities), and to be essentially verifiable, those notions must first become rationalised principles.

Historically, Heraclitus has been far from unique in his epistemological stance, which was later adopted by many Ancient schools of thought. For example, we see this in Galen's report of Democritus' theory [166]:

> If someone cannot even make a start except from something evident, how can he be relied on when he attacks his very starting-point? Democritus was aware of this; when he was attacking the appearances with the words 'By convention colour, by convention sweet, by convention bitter, but in reality atoms and void' he made the senses reply to thought as follows: 'Wretched mind, you get your evidence from us, and yet you overthrow us? The overthrow is a fall for you'.

to which the same commentator elaborates further elsewhere [166]:

> By the expression 'by convention' he means 'conventionally' and 'relative to us,' not according to the nature of things themselves, which he calls by contrast 'reality,' forming the term from 'real' which means 'true.'

According to Diogenes Laërtius, Democritus is also to have said that 'In reality we know nothing, for truth is in the depths' [166]; but we know that Democritus did apply reason to the sensible data, in his life's attempt to discover what he could, of that deep truth, with greater tenacity than any who had come before him; for although he realised how elusive it was, all accounts tell that he spent all of his inheritance travelling the world in search of knowledge, and, even, according to Eusebius, that he was reported to have said that he 'would rather discover a single explanation than acquire the kingdom of the Persians' [166]; and Aristotle—though his beliefs and his theory of Nature had been quite different—praised Democritus' use of this method, in *On Generation and Corruption* [4]: 'In general, no one except Democritus has applied himself to any of these matters [sc., coming-to-be and passing-away] in a more than superficial way. Democritus, however, does seem not only to have thought about all the problems, but also to be distinguished from the outset by his method.' A little further down, he writes that

> ... those who dwell in intimate association with nature and its phenomena are more able to lay down principles such as to admit of a wide and coherent development; while those whom devotion to abstract discussions has rendered unobservant of the facts are too ready to dogmatize on the basis of a few observations. The rival treatments of the subject now before us will serve to illustrate how great is the difference between a scientific and a dialectical method of inquiry. For, whereas the one school argues that there must be atomic magnitudes because otherwise The Triangle will be more than one, Democritus would appear to have been convinced by arguments appropriate to the subject, i.e. drawn from the science of nature;

and, eventually, when he describes the atomic theory, he begins by saying that 'The most systematic theory ..., and one that applied to all bodies, was advanced by Leucippus and Democritus: and, in maintaining it, they took as their starting-point what naturally comes first.' So we see that Democritus, and also Aristotle, as we'll see in greater detail in § 2.4, saw great worth in the practice of this Heraclitean form of natural epistemology.

It should also be admitted, however, that Sextus proceeds directly from his bastardisation of Heraclitus' philosophy, in [163], to similarly misconstrue the meaning in Democritus' writing, according to a few selected quotations; e.g., he quotes a similar passage as Galen has, that 'By convention ... ', along with similar other ones, but leaves out the senses' reply. Thus, it appears he either completely mistook, or deliberately misconstrued the meaning that had been expressed by both of these men: that primary conceptions are untrustworthy accounts of True perceptions, which can only be correctly understood through subsequent reasoning, but are nevertheless patently essential to the process of true knowledge acquisition. In other words, although, as Sextus tells us, both Heraclitus and Democritus had written that sense-perception and reason are our two inherent means of ascertaining Natural Truths (which they both believed in), and both considered sense-perception to be untrustworthy, whereas reason they thought to be reliable, the point that both had obviously tried to make, really comes down to the fact that perceptions themselves are our basic data of observation, and that biased interpretations made by the subconscious are untrustworthy, so that subsequent reasoning is necessary to form a better understanding of what those data Really are.

Probably the clearest statement of the use of this method in the development of any Ancient theory, however, which no Sceptic could deny, is given in the surviving writings of the Epicureans, who resurrected the atomic theory, in Athens, roughly fifteen years after Aristotle's death. Although most of these works were lost, a few important summaries were preserved by Diogenes Laërtius. In particular, Epicurus' letter to Herodotus [167] (X. 35-83)—his surviving epistle on physics—which was prepared as an 'epitome and manual of the doctrines as a whole', now exists, through Diogenes, in its entirety. There, Epicurus states that 'we must by all means stick to our sensations, that is, simply to the present impressions whether of the mind or of any criterion whatever, and similarly to our actual feelings, in order that we may have the means of determining that which needs confirmation and that which is obscure'; for, 'it is upon sensation that reason must rely when it attempts to infer the unknown from the known.'

As well, Diogenes preserved a list of forty 'Sovran Maxims', composed by the Epicureans from the writings of Epicurus, of which three are relevant here [167]:

22. 'We must take into account as the end all that really exists and all clear evidence of sense to which we refer our opinions; for otherwise everything will be full of uncertainty and confusion.'

23. 'If you fight against all your sensations, you will have no standard to which to refer to, and thus no means of judging even those judgements which you pronounce false.'

24. 'If you reject absolutely any single sensation without stopping to discriminate with respect to that which awaits confirmation between matter of opinion and that which is already present, whether in sensation or in feelings or in any presentative perception of the mind, you will throw into confusion even the rest of your sensations by your groundless belief and so you will be rejecting the standard of truth altogether. If in your ideas based upon opinion you hastily affirm as true all that awaits confirmation as well as that which does not, you will not escape error, as you will be maintaining complete ambiguity whenever it is a case of judging between right and wrong opinion.'

Or, according a recent gloss of the Epicurean epistemology by Robert Hicks [168]: 'It is only through sense that we come into contact with reality; hence all our sensations are witnesses to reality. The senses cannot be deceived. There can be no such thing, properly speaking, as sense-illusion or hallucination. The mistake lies in the misinterpretation of our sensations. What we suppose that we perceive is too often our own mental presupposition, our own over-hasty inference from what we actually do perceive. .... Sensations themselves must be scrutinised, and the element which the mind itself has added must be removed before we get back to the original data, the perceptions which put us in touch with reality.'

Similarly, the Stoic epistemology, which was related by Diogenes in his account of their philosophy, is thought to have risen from the foundation laid by Heraclitus. Basically, their goal was to form a consistent understanding of the grand spectrum of things by understanding 'presentations', or mental impressions, which are imprints on the soul, through subsequent thought [167]. As Diogenes related for us from an earlier text [167] (VII. 49),

> "The Stoics agree to put in the forefront the doctrine of presentation and sensation, inasmuch as the standard by which the truth of things is tested is

generically a presentation, and again the theory of assent and that of apprehension and thought, which precedes all the rest, cannot be stated apart from presentation. For presentation comes first; then thought, which is capable of expressing itself, puts into the form of a proposition that which the subject receives from a presentation."

Stoicism was founded by Zeno of Citium, also in Athens, shortly after Epicurus founded his school there, and its doctrine was widely followed for more than eight hundred years. Regarding its founder, Hicks writes [168], 'To Zeno . . . , the natural was the rational and the first mark of reason was self-consistency. A rational life must follow a single harmonious plan, whereas the paths of folly are many and various, but always stamped with inconsistency and contradiction. In this postulate of a rational law the Stoics had a precursor in Heraclitus.'

2.2.2 His Natural Philosophy

So much for Heraclitus' philosophy of knowledge and its ancient legacy: let us now turn to his philosophy of Nature—or, rather, a specific aspect of it which is the most important result of his having directly applied this method in order to try to understand Nature, which was indeed an essential step towards the present theory of Time. To be clear, most of what is known about Heraclitus' natural philosophy has to do with his mentioning of the unity of contraries, and he held that the primary element of Nature was fire. It is not immediately relevant to delve into any of this but to say that it is reasonable to admit, as, in my opinion, Barnes' argument [165] sufficiently shows, that Heraclitus' thesis was based on the principle that Nature, as a whole, is uniformly and continuously in flux—viz., in Time.—And the first explicit recognition of this principle, among the natural philosophers, lies with Heraclitus; for, if we are to take Barnes' word [165], even

> Some of those scholars who accept the Theory as Heraclitean are inclined to see nothing very original in it: the Milesians, after all, had held a similar view. The Milesians, like all observant men before Parmenides, had indeed noticed that things change: the world is patently not a static *tableau*. Yet it is far from a patent truth that *everything* changes, still less that everything *always* changes; and the Milesians, like ordinary men before Heraclitus, seem to have thought that within the changing world there was room for a number of stable courses, and the earth does not move from its place. There is no reason to deny Heraclitus the novelty of generalizing the natural view of a changing world to the more pugnacious thesis that everything changes. . .

Indeed, the common notion that everything progresses through time cannot—by definition!—be attributed to Heraclitus; e.g., we read in Plato's *Cratylus* (402a-b) [169],

> SOCRATES: I seem to see Heraclitus spouting some ancient bits of wisdom that Homer also tells us—wisdom as old as the days of Cronus and Rhea.
>
> HERMOGENES: What are you referring to?
>
> SOCRATES: Heraclitus says somewhere that "everything gives way and nothing stands fast," and, likening the things that are to the flowing (*rhoē*) of a river, he says that "you cannot step into the same river twice."

Hermogenes: So he does.

Socrates: Well, then, don't you think that whoever gave the names 'Rhea' and 'Cronus' to the ancestors of the other gods understood things in the same way as Heraclitus? Or do you think he gave them both the names of streams (*rheumata*) merely by chance?* Similarly, Homer speaks of 'Ocean, origin of the gods, and their mother Tethys'; I think Hesiod says much the same. Orpheus, too, says somewhere that 'Fair-flowing Ocean was the first to marry, and he wedded his sister, the daughter of his mother.' See how they agree with each other, and how they all lean towards the doctrines of Heraclitus.

But if we understand Heraclitus' epistemology correctly, we must recognise the significance of his own achievement of rational insight, to go beyond that which everyone before him knew, and clearly realise that *a priori* everything continually progresses uniformly through Time. Although the subconscious notion, we say, must come from the very essence of Nature, and has therefore always been, and must be, common among sentient beings, it is truly fair among humans—do with it what we will—to credit the first clear recognition of this fact to Heraclitus.

However, we must do full justice to the true brilliance of Heraclitus' insight, because the common belief among modern philosophers is that Plato did not. Namely, it is now understood that the river fragments, in which Heraclitus' principle is considered most apparent, are not supposed to be thought of as analogical or metaphorical—i.e., Heraclitus is not likening the change that occurs, as Time continuously progresses, to the flow of a river, as Plato says; nor is he even using the flow of the river symbolically, as a representation of the flow of the Universe through Time—but only as an example which clearly illustrates the uniform flux of all matter that exists throughout the Universe, regardless of its state of motion.

Indeed, consider three fragments [164]:

84a. While changing it rests.

30. <The ordered?> world, the same for all, no god or man made, but it always was, is, and will be, an everliving fire, being kindled in measures and being put out in measures.

12. As they step into the same rivers, different and <still> different waters flow upon them.

Only a further brief note to accompany each of these will be necessary:—in fragment 84a, the awkward ordering of the two verbs impresses the primacy of change, even in the thing that rests, as rest requires existence as much as any other state of being, and therefore an interval of change;—some critics have argued that the literal meanings of the surviving Heraclitean fragments can all be construed so as to admit no evidence of a principle of flux—but any critic who has ever considered that any part of Nature, according to Heraclitus' theory, might not be in continuous flux, can surely never have seen a fire—a singular thing that is never the same anywhere, from one moment to the next;—there are actually three river fragments, but

*'Rhea' sounds a lot like '*rheuma*' ('stream'); apparently Socrates expects Hermogenes to hear 'Cronus' as connected with '*krounos*' ('spring'). [Translator's note.]

the one given here is considered to be the primary one; it must be interpreted as a singlular statement regarding two different things: viz., that in one sense, humans and rivers always remain the same things; but, in another sense, that both of them continuously and uniformly change, so that although they always exist together, both do change from one instant to the next; and further, this is implicit in the fact that they *remain* the same things, which is yet another state of being.

Perhaps one of the clearest explanations of Heraclitus' principle, which nevertheless begins with the common connection to flowing water, using it metaphorically, although eventually placing another of the river fragments in context appropriately, comes from Seneca the Younger's *Ad Lucilium Epistulae Morales* (LXIII, 23) [170]:

> Our bodies are hurried along like flowing waters; every visible object accompanies time in its flight; of the things which we see, nothing is fixed. Even I . . . , as I comment on this change, am changed myself. This is just what Heraclitus says: "We go down twice into the same river, and yet [each time] into a different river."[5] For the stream still keeps the name, but the water has already flowed past. Of course this is much more evident in rivers than in human beings. Still, we mortals are also carried past in no less speedy a course; . . . every instant means the death of our previous condition. . . . So much for a man,—a substance that follows away and falls, exposed to every influence; but the universe, too, immortal and enduring as it is, changes and never remains the same. For though it has within itself all that it has had, it has it in a different way from that in which it has had it; it keeps changing its arrangement.

Thus, according to Seneca's interpretation of Heraclitus' statement, it is not the flow of the river that is to be thought of as being similar to time, but that the idea of a flowing river serves to illustrate more clearly that which is the same thing implicit in humans and rivers, as it is in all of Nature, but which is less obvious in an inert body,—that Time continuously proceeds, so that every instant is essentially different from all others—every instant in the continuous stream of Time means the death of all the Universe's previous condition, and the birth of a new one, in which the Universe nevertheless remains the unity of all things material.

Heraclitus was therefore the first true relativist:—who first realised that no rest-frame is Truly static;—who realised, before all later cosmologists, that there is a Universal rest-frame—a mean to which all the heavenly bodies and the flowing rivers within them, exist in relative motion—yet even the bodies that remain at rest with respect to that frame continually change, as the Universe changes—though while changing, It rests.

[5]'In idem flumen bis descendimus et non descendimus';—more literally, 'We twice descend and do not descend into the same river.' The original Latin is therefore more ambiguous than this translation suggests, as it *may* be taken to mean that the two existences apply singularly to the act of descent, rather than to the different states of the physical system. However, given the context, we can be sure that *Seneca's* meaning has been translated correctly, as he probably did not mean to state something that has nothing whatsoever to do with his line of thought. I have therefore added the words 'each time', in order to give further clarification to this interpretation.

2.3 On Eleatism

As we now move from Heraclitus', to a description of the Eleatic theory, it is worth noting, first of all, that, whereas Heraclitus' theory was based on the reduction of observational evidence to basic principles, the Eleatics based theirs on systemmatic logical deduction. Now, our primary reason for analysing the Eleatic theory, is so that we may investigate an extremely serious philosophical problem that was first propounded by Parmenides, which remains a significant problem in understanding the meaning of relativity theory; but because it has general relevance in our present endeavours, and so that we may clearly understand their theory—in particular, the origin of this persisting problem,—we also have cause to further investigate their method of inquiry.

The best way to understand the fragmentary statement of Parmenides' theory is to begin with a clear description of the basic principle of his logical deduction. The reason for this, is that in its fragmentary form Parmenides' poem is probably too ambiguous for a conclusive literary deduction, but it seems that with this prior understanding, its meaning immediately becomes quite clear, as the reasoning behind each line can be more easily grasped.

The general idea is that what appears to be a certain conclusion that could be taken from Parmenides' argument, is actually the prime motivation for his entire theory of nature, which results from a straightforward deduction based on a single principle that he decided to assume, in all likelihood incorrectly, and which is corroborated by the resulting theory's partial resolution of the problem of becoming. More specifically, consider a younger Parmenides, thinking about Nature—about being, becoming, and change:—he reasons that Being Is and Not-being Isn't, and nothing which Isn't can actually exist. If Change Is,—from what, and where to? If from, and into, That-which-does-not-presently-exist, then surely Change cannot Be, because past-state could not thereby become What-is, nor can present-state become What-is-not. Therefore, the past and future, which common sense tells us were, and will be, must Be *now*, as with the present, and time must be an illusion. Therefore, the One that is All must Be all of space and time, absolutely determined in all four dimensions, in a singular temporal sense, as a Block.

The remainder of the theory, which is brilliant in its own right, is built upon this; I'll remark only on one aspect which should be of interest to the modern relativist: viz. that Parmenides realised,—as Einstein also would, two and a half millennia later, for his own particular reasons,—that there would be a problem with his cosmology if space would be infinite; so he did *precisely* what Einstein did, and made it spherical. However, so that it would have no beginning or end in time, he made the One (what can *now* only be recognised as) a 4-sphere (although his theory was not formally so explicit), which he explained by the only conceptually possible analogy (in any millennium) of an embedded 2-sphere.

Now, what should become plain through consideration of the above interpretation of what Parmenides' logic *seems* to have been, is that his entire theory hinges on one basic principle: that 'time' and 'space' have no real existence in and of themselves, independent of the material that Is there; i.e., that the One is *purely* material, and there is no 'where' for anything to Be—that Being is *material* and Not-being is *immaterial*, and *matter* is all that there Is. As Giorgio de Santillana says [171], the concept of Being in Parmenides' poem is pure geometrical space, and '[g]eometry as the Greeks meant it put three requirements on its space: first, it must have continuity (in a sense somewhat stronger than the mere absense of gaps between points); second, it must be the same, homogeneous throughout, so that we can move figures freely from place to place without altering their geometrical properties; and, finally, it must be isotropic, or the same in all directions'; furthermore, it is a non-trivial fact that he does actually formulate this as a purely *corporeal* existence, with matter and existence being perfectly synonymous things, and all that there Is, so that the concept of 'place', e.g., as defined later by Aristotle, in his *Physics* [4], is purely a delusion; i.e., that the Reality that Is is to be a purely physical Corporeality, existing as an isotropic, homogeneous, complete metric space, though somehow also abstract to the senses.

If this basic principle had actually been correct, it would be much more difficult to disagree with Parmenides' theory than we shall presently find, as Plato and Aristotle both realised, who both believed that nothing physical could be incorporeal, which problem they circumvented in proposing the reality of abstract, nonphysical things. This problem, first raised by the Eleatics, is of paramount importance in metaphysics; therefore, we must take a brief pause, in order to examine the various forms of real existence that were proposed as a direct result of the Eleatic reasoning, so that we may better understand both the choice they made and, eventually, our own metaphysical theory.

We therefore begin by distinguishing the classifications of corporeal and incorporeal, and physical and nonphysical, as defining absolute complements of two different forms of real existence. In modern physics, e.g., we say that the elementary particles and the compounds formed from them are the corporeal things, whereas the vacuum of spacetime is incorporeal; and we say that anything—i.e., any body, process, state, etc.—that is subject to the fundamental Laws of physics, whatever they may be, is physical, whereas anything else is nonphysical. In fact, it remains a possibility that all nonphysical things might have physical counterparts; e.g., that an idea should accompany a certain physical situation and occurrence of synapses in a brain. However, it is not our purpose here to judge whether any nonphysical thing could actually be real, in and of itself, or whether all must be byproducts as such; rather, in our later theoretical development, we shall stick to the forms of physical existence, with the aim of giving a sufficient self-consistent description of them.

But before we can assume that, we must see how the various theories of real existence generally apply in theories of Nature. First of all, it may be reasonably posited that all things corporeal must be physical; so we are led to identify three possible forms of real existence: physical corporeal, physical incorporeal, and nonphysical incorporeal. Then, as I've said in so many words, the Eleatics believed, and this was the principal tenet of their theory, that the only real existence could be a physical corporeal continuous geometry; to them, whether a body's *place* was a physical incorporeal container, or a useful concept for describing things in common speach, as they exist in the mind, did not matter because neither would actually be real. And if 'space' and 'time' are not real things, then what Is could not actually travel through them—it could not come from the past, or move into the future, any more than it could move from place to place.

Following the Eleatic line of reasoning, but not their ultimate beliefs, the Atomists, Leucippus and Democritus, posited that a physical and incorporeal void should also be real, though they continued to deny the actual existence of anything nonphysical. In their theory, the void would be continuous and would contain matter composed of finite indivisible atoms. Conversely, Plato took up the alternative,—that reality consists of the physical corporeal and the nonphysical incorporeal,—and Aristotle further developed the physical theory that would naturally describe this, i.e., eventually arriving at an abstract physical description of time, in which the material Universe would be continually rearranged; for so he found that matter should be everywhere, but that the places containing matter could also be real, and change as well; and that acting by mutual contact, 'some things will suffer action and others will act, provided they are by nature adapted for reciprocal action and passion' [4]. As we'll see in § 2.6, it was St. Augustine who eventually provided the sufficient argument that favours Time as a physical thing, though he did not himself believe this.

Thus, I think—and this is precisely how the problem was taken up later by the Atomists, whose methods in this respect Aristotle placed above all others, as mentioned above—that the problem with Parmenides' thesis is merely that he started off on the wrong foot: he began from a basic principle he believed was reasonable to assume, then used further reason and reached an inconsistency; but rather than re-tracing his steps and objectively inspecting his basic principles, he forced himself to adjust his understanding of Nature, describing it as an abstract and unverifiable thing, in an attempt to reconcile the issues he had found—and he was probably happy to do so, as he found Genesis to be less of a problem. Personally, if I wished to expound such a theory with artful craft, I should hope to have written something like this [172]:

> Come, I shall tell you, and do you listen and convey the story,
> What routes of inquiry alone there are for thinking:
> The one—that [*it*] *is*, and that [*it*] *cannot not be*,
> Is the path of Persuasion (for it attends upon truth);
> The other—that [*it*] *is not* and that [*it*] *needs must not be*,
> That I point out to you to be a path wholly unlearnable,
> For you could not know what-is-not (for that is not feasable),
> Nor could you point it out.
> . . .
> It must be that what is there for speaking and thinking of *is*; for [it] is there to be,
> Whereas nothing is not; that is what I bid you consider,
> For <I restrain> you from that first route of inquiry,
> And then also from this one, on which mortals knowing nothing
> Wander, two-headed; for helplessness in their
> Breasts guides their distracted mind; and they are carried
> Deaf and blind alike, dazed, uncritical tribes,
> By whom being and not-being have been thought both the same
> And not the same; and the path of all is backward-turning.
> . . .
> For never shall this prevail, that things that are not *are*;
> But do you restrain your thought from this route of inquiry,
> Nor let habit force you, along this route of much-experience,
> To ply an aimless eye and ringing ear
> And tongue; but judge by reasoning the very contentious disproof
> That has been uttered by me.
> . . .
>   A single story of a route still

Is left: that [*it*] *is*; on this [route] there are signs
Very numerous: that what-is is ungenerated and imperishable;
Whole, single-limbed, steadfast, and complete;
Nor was [it] once, nor will [it] be, since [it] is, now, all together,
One, continuous; for what coming-to-be of it will you seek?
In what way, whence, did [it] grow? Neither from what-is-not shall I allow
You to say or think; For it is not to be said or thought
That [*it*] *is not*. And what need could have impelled it to grow
Later or sooner, if it began from nothing?
Thus [it] must either be completely or not at all.
Nor will the strength of trust ever allow anything to come-to-be from what-is
Besides it; therefore neither [its] coming-to-be
Nor [its] perishing has Justice allowed, relaxing her shackles,
But she holds [it] fast; the decision about these matters depends on this:
Is [*it*] *or is* [*it*] *not*? but it has been decided, as is necessary,
To let go the one as unthinkable, unnameable (for it is no true
Route), but to allow the other, so that it is, and is true.
And how could what-is be in the future; and how could [it] come-to-be?
For if [it] came-to-be, [it] is not, nor [is it] if at some time [it] is going to be.
Thus, coming-to-be is extinguished and perishing not to be heard of.
Nor is [it] divisible, since [it] all alike *is*;
Nor is [it] somewhat more here, which would keep it from holding together,
Nor is [it] somewhat less, but [it] is all full of what-is.
Therefore [it] is all continuous; for what-is is in contact with what-is.
Moreover, changeless in the limits of great chains
[It] is un-beginning and unceasing, since coming-to-be and perishing
Have been driven far off, and true trust has thrust them out.
Remaining the same and in the same, [it] lies by itself
And remains thus firmly in place; for strong Necessity
Holds [it] fast in chains of a limit, which fences it about.
Wherefore it is not right for what-is to be incomplete;
For [it] is not lacking; but if [it] were, [it] would lack everything.
The same thing is for thinking and [is] that there is thought;
For not without what-is, on which [it] depends, having been declared,
Will you find thinking; for nothing else <either> is or will be
Besides what-is, since it was just this that Fate did shackle
To be whole and changeless; wherefore it has been named all things
That mortals have established, trusting them to be true,
To come-to-be and to perish, to be and not to be,
And to shift place and to exchange bright colour.
Since, then, there is a furthest limit, [it] is completed,
From every direction like the bulk of a well-rounded sphere,
Everywhere from the centre equally matched; for [it] must not be any larger
Or any smaller here or there;
For neither is there what-is-not, which could stop it from reaching

[Its] like; nor is there a way in which what-is could be
More here and less there, since [it] all inviolably *is*;
For equal to itself from every direction, [it] lies uniformly within its limits.
Here I stop my trustworthy speech to you and thought
About truth; from here onwards learn mortal beliefs,
Listening to the deceitful ordering of my words ...

As Barnes points out [165], Parmenides' doctrine is not based on an epistemological principle that common sense must be denied in objective philosophical deduction,—that he does not say, 'don't listen merely to your eyes and ears, and bear witness to others' testimonies which are based on sense-perception'; but says 'listen to reason, and if reason produces a result that is objectionable to common sense, do not deny it without a reasonable rebuttal'.

This is a reasonable enough assertion; but it is equally reasonable to say that if some observation of Nature poses a problem to a theory that has been deduced from a certain set of facts, the worst kind of cop-out for the natural philosopher is to immediately assume that thing is not really a part of Nature—that Nature must be otherwise—and then attempt to construct a philosophy based on that assumption, without first objectively examining the principles which led to the original problem—for many more shall inevitably follow. Here, Parmenides seems to have disposed of two great philosophical problems by denying the common sense notions of change, alteration, flux, motion, etc.—but only provisionally, as he has done this without having adequately addressed the more critical problem of reconciling his theory with the empirical evidence.

Here, I say '*seems* to have disposed', because I wonder,—has he really done away with the problem of the origin of Nature by assuming all four dimensions are really a temporally singular entity? He has not explained the origin of that singularity, but he argues correctly that this mode of inquiry is a misunderstanding of his thesis. So we ask instead, Why *is* reality the One? To which we find only the entirely unacceptable anthropic or divine responses. I concede that his may be closer to a solution than, say, spontaneous generation, which is met with the multiple snags well-known in Big Bang cosmology; but it is no better, in this regard, than what the Atomists later came up with, in response to Parmenides' argument. So we should in fact say that he has *circumvented* the problems of the Universe's coming-to-be, and why It is as it Is.

There is an important contrast that can be made with Galileo's achievement in discovering the relativity of inertia: i.e., that when one principle, which is necessarily based on some empirical evidence, appears to be in conflict with some other known datum, it cannot do to simply deny either of these, rather than examining all of the data objectively in order to find a principle consistent with all, as it could well be the principle in conflict that is incorrect; i.e., that the basic principles used in the deduction of a theory must be reduced when that theory has become liable—and it is not surprising that this was indeed the problem with the first theory ever logically deduced.

But in Galileo's case, the principle that the Earth is at the centre of the Universe was based on the fact that we immediately infer this from basic perception, and had therefore been part of many theories since Ancient times—though not all of them. However, observations of the motions of celestial bodies, and the principle of a True physical order in Nature, could not be reconciled with the geocentric principle, which similarly led Kepler to the discovery of his Three Laws—although Galileo, who probably didn't know of Kepler's Laws [173], for it seems that Kepler himself never correctly understood the worth of his *physically meaningful* discoveries, and therefore never supported them as he did his more fantastic endeavours [174], based his argument for heliocentrism primarily on evidence gained from telescopic observations [175]. Therefore, the problem was met by two conflicting principles, each supported by different empirical evidence, and the problem became to determine which was false. By reasoning that we would not actually notice any difference if the Earth were in a heliocentric orbit, due to the relativity of inertia, which is also empirically supported, and shows that the common sense-based geocentric principle was basically flawed, Galileo was able to advance an intuitive argument that was again supported by empirical evidence.

Conversely, Parmenides gave no empirical evidence to support his claim that we are delusional about change and motion; his argument was that this should have been accepted because it was the result of a more substantial, logically deduced theory, and that the need for general consistency should be trumped by such pure reason, even if that should neglect some relevant facts. In fact, the very resolution to his problem lies in that which he has led himself to deny: for if we move through space and time, as common sense tells us we do, then space and time should be ontologically real, regardless of whether something is actually there. The logical implication is that there *could be* places in space or time where no material is present,—that spacetime is not a continuous material *plenum*,—which would then allow for the motion of matter through those void regions. Although the Eleatics apparently did realise this, they abhorred the prospect that Not-being could Be, and thus failed to accept the possibility of separating the incorporeal from the corporeal within the realm of physical existence—and this is precisely what Leucippus and Democritus *did* do.

However, it is obvious, and has always been clearly understood, that the Eleatic doctrine cannot be refuted scientifically; for nothing can be ultimately refuted if it cannot possibly be detected, and this is indeed the case with the Eleatic hypothesis that there is one singular eternity. Science inductively infers hypotheses and verifies their correctness through as many different repeatable observations as possible, or else falsifies them by a single one; therefore, an hypothesis that is impossible to verify *once* is not scientific. But one which is totally at odds with the data we do possess, however untrustworthy we might think they are, is also physically useless. The truth is, that an intuitive example that would suggest Heraclitus' principle of Universal flux—the first principle of relativity—is incorrect, has never been found in Nature—and therein lies the best scientific evidence we can ever hope to find, that it is.

Generally speaking, without a sufficient reductive argument to provide a new, but nevertheless equally acceptable, intuitive understanding of Nature, any theory that may be logically deduced, which finds itself in stark conflict with the appearance of Reality, and the principles that may be directly inferred from that, cannot be allowed to replace such natural intuition; its merits may be weighed, and, even if it is found to have predictive power beyond that of any other contending theory, it shall never be an acceptable theory of Nature until sufficient principle reduction has been achieved.—As Russell said before [176],

> I cannot believe—and I say this with all the emphasis of which I am capable—that there can ever be any good excuse for refusing to face the evidence in favour of something unwelcome. It is not by delusion, however exalted, that mankind

can prosper, but only by unswerving courage in the pursuit of truth.

With good reason, Barnes calls the part of his book dealing with the unfalsifiable Eleatic theory 'The Serpent' [165], as its Devilish charm has seduced lazy and obtuse metaphysicists ever since. We find, however, in full examination of the situation, that Heraclitus' principle must be included in a truly objective deduction of the known facts; and, if neither theory is favoured when all others have been considered, we say that this is indeed the critical point—for it manages to tip the scale vertically, in favour of the Heraclitean theory. For this reason, the Eleatic doctrine has never truly become the paradigm in natural science, which seems often to find itself taking a Sceptical stance when faced with such theoretical inconsistencies as the Eleatics discovered.

2.4 Accession to a Natural Epistemology

In everything that has now been said against the Eleatic doctrine, we find ourselves in the company of good authority, as the main points made here were similarly stated by Aristotle at the beginning of his *Physics*, as he, too, argued that the physical description of Nature must be based on natural principles [4] (184b26-185a14):[6]

> Now to investigate whether what exists is one and motionless is not a contribution to the science of nature. For just as the geometer has nothing more to say to one who denies the principles of his science—this being a question for a different science or for one common to all—so a man investigating *principles* cannot argue with one who denies their existence. For if what exists is just one, and one in the way mentioned, there is a principle no longer, since a principle must be the principle *of* some thing or things.
>
> To inquire therefore whether what exists is one in this sense would be like arguing against any other position maintained for the sake of argument (such as the Heraclitean thesis,[7] or such a thesis as that which exists is one man) or like refuting a merely contentious argument—a description which applies to the arguments both of Melissus and of Parmenides: their premisses are false and their conclusions do not follow. Or rather the argument of Melissus is gross and offers no difficulty at all: accept one ridiculous proposition and the rest follows—a simple enough proceeding.
>
> We, on the other hand, must take for granted that the things that exist by nature are, either all or some of them, in motion—which is indeed made plain by induction. Moreover, no one is bound to solve every kind of difficulty that

[6]See also Aristotle's *Physics* 253a32-b6 [4].

[7]Presumably, as mentioned further on (185b19-25) [4]: 'But if all things are one in the sense of having the same definition, like raiment and dress, then it turns out that [Melissus and Parmenides] are maintaining the Heraclitean doctrine, for it will be the same thing to be good and to be bad, and to be good and to be not good; and so the same thing will be good and not good, and man and horse . . . .' We can assume from this statement, that Aristotle saw no epistemological merit in Heraclitus' rant about the unity of contraries, and therefore gave no recognition, as we have above, that his method of Natural inquiry had anticipated the one that is currently being argued for.

> may be raised, but only as many as are drawn falsely from the principles of the science: it is not our business to refute those that do not arise in this way . . .

The everlasting impression of an idea that can never be refuted, and a Devilish seduction that has always found a way to draw in some Idealist sympathisers,—which has been thought by some Sceptics to be as much correct as it isn't,—is one way of looking at the Eleatic legacy; but a fairer assessment would be that the Eleatics had a logically self-consistent theory, however immature, and that they supported it with a good argument against Scepticism.

For, as the Eleatics reasoned, the way to truly refute a theory is not to criticise its inconsistency with traditional opinion, but to fully examine the arguments for and against it, and decide, based on that, which is the most reasonable position to maintain. This is the main argument that is disseminated in Plato's *Parmenides*, and the point was in fact explicated by Parmenides in his poem, which was written in two such parts: *The Way of Truth* (quoted above) and *The Way of Opinion* (which is mostly lost).

Although the Sceptics also have a reasonable point, which shall be propounded here as well, that absolute Truths cannot be known with absolute certainty, and a theory's overall consistency with empirical evidence must be ultimately required, a Sceptic contributes nothing meaningful to our understanding of Nature with the paltry argument, that a theory cannot be correct if it is inconsistent with this premiss—and, in fact, often represses the search for a more correct theory by voicing doubts which are based on nothing more than a difference of opinion, rather than attempting to sort out what the crux of that difference really is. Accordingly, many academics in the sixteenth and seventeenth centuries were unwavering in their adherence to the geocentric principle.

Conversely, although we shall remain consistent with some basic principles of Scepticism, the aim of the present argument shall be more like that of the Eleatics; for, rather than allowing the inner Sceptic to merely voice obvious criticisms, the primary aim here has been to clearly *understand*, through logical analysis, the precise reasons why very brilliant people allowed themselves to accept theories of Nature which are obviously incorrect; for we would be at the height of ignorance to presume that geniuses like Parmenides and Einstein did not have very good reasons for propounding the theory that time is merely an illusory aspect of a four-dimensional 'Block'. And it will be clearly shown how wrong it has been to ignore the fact that this *is* what their theories formally required, and go on using them to explain processes which are patently dynamical; for, by analysing both sides of the argument in the Eleatic tradition, we shall be able to determine where the true points of conflict lie, and see clearly that the principles which end up formally requiring a strictly determined, substantive block, are precisely the ones that have also led to significant mental blocks, as paradoxes have been subsequently inferred when attempts were made to describe dynamical processes by the same theories.

So we acknowledge the significant contribution of the Eleatics, who developed the method of using deductive reasoning in order to derive a logical theory that is consistent with basic principles. Once any such theory exists, it is possible to use scientific induction, to state testable hypotheses which allow us to refine our knowledge of Nature as important discoveries happen to be made. As Thomas Kuhn explained, however, this process of refinement is not continuous, as, along with the discoveries which come in line with the current paradigm, significant problems are also continually discovered, which eventually lead to paradigm shifts, when a new theory is discovered which brings that knowledge all into line. To quote the blurb from his seminal essay on *The Structure of Scientific Revolutions* [177],

> Drawing his data from history, philosophy, and psychology, Kuhn argues that "normal science" presupposes a conceptual and instrumental framework or paradigm accepted by an entire scientific community; that the resulting mode of scientific practice inevitably evokes "crises" which cannot be resolved within this framework; and that science returns to normal only when the community accepts a new conceptual structure which can again govern its search for novel facts and for more refined theories.

It is, in fact, reasonable to suggest that these paradigm shifts might ultimately arise from discoveries made according to the less commonly practised Heraclitean method of introspection, or principle reduction, which is a necessary element of the method of inquiry required to form a complete understanding of the workings of Nature; consequently, it would seem that only through the marriage, and continued practice, of these two essential methods,—in probing and discerning the potentially basic principles of Nature, and then deducing greater theories in agreement with the novel facts of observation, which ultimately leads to further testable hypotheses,—does the paradigm of natural philosophy truly progress—i.e., that by systemmatically probing the limits of logically deduced, generally consistent metaphysical theories by both induction and reduction, we form a clear and self-consistent idea of what Nature really is.

In truth, though, Heraclitean reduction is often neglected in practice, whenever the current paradigm is assumed as an absolute Truth. Then, it seems that the paradigm itself eventually becomes foggy, due to a common neglect of the necessary continuous introspective rationalisation, and then unclearly defined, due to a subsequent necessity to reconcile the paradoxes that ultimately result when inductive progress is not matched by the necessary reduction, or a general deduction has been neglected when parts of the theory have become too specialised and the consistency of the whole has thus become difficult to infer; and science allows itself to sumble about, half-blind, through an interval bounded at the one end by an unsure, yet comfortably assumed bias, and at the other, by the boundary that assumption places on the possibilities of theoretical advancement—for the two ends are really parts of a continuous interval, of which progress can only go so far in one direction, without seeing an equivalent amount in the other.

Thus, the historical evidence has shown that after any paradigm has enjoyed the comfort of its respite, and long tradition has replaced original reasoning, as the agent of its acceptance—after blind stumbling has eventually allowed progress to be made in one direction only, having uncovered as many novel facts and theoretical refinements as possible, yet these are accompanied by an equivalent amount of problems—we see that the most sure way to progress when the continued practice of scientific induction alone has led to the realisation of significant problems or inconsistencies within the current paradigm, is *not* to continue piling *a posteriori* hypotheses and conjectures onto that paradigm at leisure, nor to propose baseless untestable new theories with naturally uninspired mathematical finesse and hopeless desparation—in both cases relying on blind luck and a remote possibility of some eventual observation, etc., for validation of such educated guesswork,—but to rationally examine the basic principles of the current paradigm, with the goal of further clarifying what they must be, and why; and possibly, with that clearer understanding, reducing them to more basic ones, and thereby emerging with a new and clearer view of Physical Reality, as well as a theory that is consistent with the old one, but which also accounts for the problems that have been realised through inductive inference from the empirical data, according to subsequent objective deduction.

To state this critical point more plainly, it is not enough to try to objectively interpret our observations, which interpretations inevitably conform to the current paradigm, even if they are thereby understood to be paradoxical, but we must also constantly, prudently reason our way to the Truth. For the ultimate goal of all this, is not merely a model deduced from the current theory which would account for the observed facts; and it is cheap, and therefore generally ineffectual, to create a whole new unverifiable theory out of nothing but imprudent philosophical speculation (even if that should be mathematical in nature) that has no prior connection, common sense-based or otherwise, to a realistic understanding of nature. The goal is in fact to understand 'Why?' And this question cannot be answered without a clearly reduced metaphysical theory based on concrete natural principles, from which such a model—i.e., one that would account for all the observed facts—should naturally emerge, according to those prior natural causes.

The age-old hitch in actually achieving this goal, as we know, is the inability or unwillingness of subsequent practitioners, to refine their basic understanding once a problem with the accepted physical theory has been realised, as they would still rather work within the bounds that are set by the paradigm it created; or, as Einstein put it in his 1916 tribute to Ernst Mach [178],

> .... Concepts which have proven useful in ordering things, easily attain such an authority over us that we forget their Earthly origins and accept them as unalterable facts. They are then branded as 'necessities of thought', 'a priori givens', etc. The path of scientific advance is often made impassable for a long time through such errors. It is therefore by no means an idle trifling, if we become practiced in analysing the long-familiar concepts, and to show upon which circumstances their justification and applicability depend, as they have grown up, individually, from the facts of experience. For through this, their all-too-great Authority will be broken. They will be removed, if they cannot be properly legitimated, corrected, if their correlation to given things was far too careless, or replaced by others, if we see a new system that can be established, that we prefer for whatever reasons.
>
> This type of analysis appears to the scholars, whose gaze is directed more at the particulars, most superfluous, splayed, and at times even ridiculous. The situation changes, however, when one of the habitually used concepts should be replaced by a sharper one, because the development of the science in question demanded it. Then, those who are faced with the fact that their own concepts do not proceed cleanly raise energetic protest and complain of revolutionary threats to their most sacred possessions. In this cry, then, mix the voices of those philosophers who believe those concepts cannot be done without, because they had them in their little treasure chest of the 'absolute', the 'a priori', or classified in just such a way that they had proclaimed the principle of immutability.

And so it is regrettable that physicists have not since made greater strides to incorporate such investigations of the ontological side of physical theory into regular practice, along with the study and development of appropriate epistemological methods, and thus work to close the gap between philosophy and science as physical theory has developed, rather than leaving the philosophical part of theoretical development to be investigated only by the philosophers whose opinions they're apt to ignore; for, as Einstein also wrote in this article, from the height of scientific achievement [178],

> But how does it happen anyway, that a properly endowed natural scientist comes to concern himself with epistemology? Is there not more valuable work in his own trade? I hear some of my colleagues saying this, or sense from many more that they feel this way. I can not share this attitude. When I think about the ablest students I have encountered in my teaching, viz. those who distinguish themselves through independence of judgment, and not through sheer agility only, so I state of them that they had a lively interest in epistemology. They gladly entered into discussions about the aims and methods of the sciences, and showed unequivocally, through persistence in advocating their views, that the subject seemed important to them. In truth, this is not surprising.
>
> If I am not ambitious for external reasons, such as making money, and also not, or at least not exclusively, the sporting pleasure, or delight in brain-gymnastics due to a scientific turning, then, as a disciple of this science, I must have a burning interest in the question: What possible goal does the science want to reach, to which I dedicate myself? To what extent are its general results 'true'? What is essential, which is based solely on accidents of development?

The significance of this last question warrants some elaboration, so that its meaning should not be missed: in common scientific practice, we inductively infer certain properties of Nature according to empirical observation, and scientific theory grows by piecing together all of these primary inferences. But it is only by stepping back from the specifics of the theory and considering the whole of scientific and philosophic knowledge objectively, while examining the circumstances upon which the justification and applicability of the long-familiar concepts depend, that we can rationally reduce those 'accidents of development' to principles which are closer to essential properties of Physical Reality.

Accordingly, the most novel of observations, such as the discovery that the cosmic expansion is now accelerating, should not be merely fitted into the scientific paradigm that exists at the time of discovery; but reason, in conjunction with mathematics, should be applied to such problems in order to account for their consistency with the whole set of knowledge.

Now, Einstein's convictions about the essential importance of being ever-conscious of the epistemological methods one uses in conducting scientific investigation was no mere passing fancy either; as he stated unequivocally in 1949 [40],

> The reciprocal relationship of epistemology and science is of noteworthy kind. They are dependent upon each other. Epistemology without contact with science becomes an empty scheme. Science without epistemology is—insofar as it is thinkable at all—primitive and muddled.

The reason for this, is simply that a physical theory which aims to accurately understand and describe Physical Reality cannot be complete without a corresponding clear notion (which is often achieved through elucidation) of its epistemological intent, without which it often loses sight of its purpose, its meaning, and its logical coherence. But in practice, the requirement of such constant contact with a coherent epistemological method is no trivial task to consciously maintain; and, as Einstein went on to explain after the above statement, there is as yet no strict system general enough to properly account for all the possibilities that arise in the process of fundamental inquiry [40]:

> However, no sooner has the epistemologist, who is seeking a clear system, fought his way through to such a system, than he is inclined to interpret the thought-content of science in the sense of his system and to reject whatever does not fit into his system. The scientist, however, cannot afford to carry his striving for epistemological systematic that far. He accepts gratefully the epistemological conceptual analysis; but the external conditions, which are set for him by the facts of experience, do not permit him to let himself be too much restricted in the construction of his conceptual world by the adherence to an epistemological system. He therefore must appear to the systematic epistemologist as a type of unscrupulous opportunist: he appears as *realist* insofar as he seeks to describe a world independent of the facts of perception; as *idealist* insofar as he looks upon the concepts and theories as the free inventions of the human spirit (not logically derivable from what is empirically given); as *positivist* insofar as he considers his concepts and theories justified *only* to the extent to which they furnish a logical representation of relations among sensory experiences. He may even appear as *Platonsit* or *Pythagorean* insofar as he considers the viewpoint of logical simplicity as an indispensable and effective tool of his research.

The goal of such investigations is indeed a formal analytical theory which is built on basic principles or axioms, using them to develop a mathematical formalism that would describe the empirical data; and the theoretical physicist must therefore be more than one endowed with a great arsenal of mathematical techniques:—he must be an analyst in the full sense, who rationally scrutinises his theory and further develops its possibilities through logical and mathematical deduction in order to naturally 'save the phenomena'; he must be a rational (foundationalist)-empirical (idealist) analyst who therefore accepts the fact that he can never presume the Correctness of his theory, which must necessarily be built upon axioms, as well as a naturalist who seeks a logically coherent, and naturally *seamless* connection between such axioms and all that can be discerned from Nature.

In his 'obituary' [40], Einstein wrote vaguely (sc., with 'meager precision') of how this should be done, when he discussed the two 'points of view according to which it is possible to criticize physical theories at all';—the one of which concerns itself with the 'premises of the theory itself, with what may briefly but vaguely be characterized as the "naturalness" or "logical simplicity" of the premises', i.e. 'with the "inner perfection" of the theory';—and the other, primary point, which 'refers to the "external confirmation"', is 'concerned with the confirmation of the theoretical foundation by the available empirical facts', he stated precisely as the requirement 'that the theory must not contradict empirical facts';—and, furthermore, that this strict adherence to observation must not be achieved through 'the adaptation of the theory to the facts by means of artificial additional assumptions'; for, in this way, by continually stressing both of these critical limits, and facilitating their seamless connection through logical deduction which is often realised through precise mathematical expression, 'we are confining ourselves to such theories whose object is the *totality* of all physical appearances'.

Therefore, the end goal of fundamental inquiry is to achieve a basic theoretical expression for our Cosmic Existence, which lends itself naturally to a description of all the observed phenomena that does not require unqualified presumption, or *ad hoc* definition, of any characteristics of Nature, such as action-at-a-distance, empirically given 'fundamental' constants, etc; for 'there are no *arbitrary* constants of this kind; that is to say, nature is so constituted that it is possible logically to lay down such strongly determined laws that within these laws only rationally completely determined constants occur (not constants, therefore, whose numerical value could be changed without destroying the theory)' [40].

A final remark must be made before we can move beyond this discussion of Einstein's epistemological methods:—it may also be that the scientist, and even one with the noblest intentions, should fall into the trap to which Einstein condemned systematic epistemologists in [40]; for when his physical theory is finally complete, he may have become too immersed in it to be able to see how to further reduce his principles, expand the theory, and therefore see how it can be brought into greater agreement with sense-perceptions; and this indeed appears to have occurred for both Einstein and the Eleatics. With Einstein, the reason for this seems most likely to have been a caveat that warrants mention in regard to the goal of logical simplicity or logical economy of principles, which he held to be the mark of a great theory, and therefore took as his aim. The problem with this, as I see it, is that when one actively seeks out principles that would satisfy this requirement, and does not require above all the 'naturalness' or 'logical seamlessness' of the theory, along with the intuitive clarity that these afford, one may be led into abstraction by neglecting to properly account for certain facts of observation in this way, as Einstein was. Along with that which was noted at the end of Chapter 1, this point seems to stand out, e.g., in a passage from 'The Problem of Space, Ether, and the Field in Physics' (1934) [179]:

> The theory of relativity is a fine example of the fundamental character of the modern development of theoretical science. The initial hypotheses become steadily more abstract and remote from experience. On the other hand, it gets nearer to the grand aim of all science, which is to cover the greatest possible number of empirical facts by logical deduction from the smallest possible number of hypotheses or axioms. Meanwhile, the train of thought leading from the axioms to the empirical facts or verifiable consequences gets steadily longer and more subtle. The theoretical scientist is compelled in an increasing degree to be guided by purely mathematical, formal considerations in search for a theory, because the physical experience of the experimenter cannot lead him up to the regions of highest abstraction. The predominantly inductive methods appropriate to the youth of science are giving place to tentative deduction. Such a theoretical structure needs to be very thoroughly elaborated before it can lead to conclusions which can be compared with experience. Here, too, the observed fact is undoubtedly the supreme arbiter; but it cannot pronounce sentence until the wide chasm sep-

> arating the axioms from their verifiable consequences has been bridged by much intense, hard thinking. The theorist has to set about this Herculean task fully aware that his efforts may only be destined to prepare the death blow to his theory. The theorist who undertakes such a labor should not be carped at as "fanciful"; on the contrary, he should be granted the right to give free reign to his fancy, for there is no other way to the goal. His is no idle daydreaming, but a search for the logically simplest possibilities and their consequences. This plea was needed in order to make the listener or reader more inclined to follow the ensuing train of ideas with attention; it is the line of thought which has led from the special to the general theory of relativity and thence to its latest offshoot, the unified field theory.

Now, I do not wish for my point to be misunderstood here: I agree fairly well with everything in this statement. For logical economy can, and indeed should, be held in sight when searching for a basic theory of Nature; but it cannot be the *ultimate* goal of the metaphysicist (for that is what one is, if one begins to rationally scrutinise the meaning of physical theory and its connection to Physical Reality, rather than merely conducting experiments based on the prior theorisation of others, with the goal of making inductive inferences to advance that theory) to derive the most economical principles with the farthest reaching consequences, if that would come with the price of abstraction due to a neglect of any reproducible empirical datum—and especially not one that was accounted for in the older theory;—for that is no proper reduction.

And the fact is, that the Newtonian concept of an absolute time, though too carelessly formulated in the mathematical theory, is in accord with Heraclitus' principle, and thus with a rationalisation of common sense-perception that is no less pertinant to our theoretical description of Physical Reality than any physical interpretation of an observation that might be made, e.g., at the Large Hadron Collider; therefore, the Newtonian theory possesses an intuitive clarity which the theory of relativity, according to the common interpretation that leads to a block universe description in no less straightforward a manner than the Eleatic deduction, does not. As Einstein wrote in his autobiography [40],

> Newton, forgive me; you found the only way which, in your age, was just about possible for a man of highest thought- and creative power. The concepts, which you created, are even today still guiding our thinking in physics, although we know that they will have to be replaced by others farther removed from the sphere of immediate experience, if we aim at a profounder understanding of relationships.

The notion that certain principles can be simply abandoned for others in order to achieve the 'grand aim of all science', is a recurring theme in Einstein's epistemology, and a mark against it. For in order to properly reduce the concepts, hypotheses, or axioms of any theory that does hold the weight of Authority, as it had previously been thought to provide a good description of Nature, *only* the second option that Einstein stated in his tribute to Mach can be allowed; viz., they must be corrected, by showing that their correlation to the given things was far too careless. For then it is reasonable to expect that logical economy shall be an accidental property of the resulting theory.

The epistemological method discussed here, is in fact the method that has been practised, to some degree or other, by all of the greatest thinkers in history who have made the most important advances in our understanding of Nature. In ancient times, we see its mark in Democritus' search for Truth, the worth of which he placed above the kingdom of the Persians, and in the general completeness of his theory, which he derived by rational thought based on primary natural observations; and in Aristotle, who not only meticulously deduced a physical theory based on such prior natural principles, that was as epistemologically complete as any has been since, but who also exhaustively formalised the requisite methods of logical inquiry; and we know this to have been a central aspect of Epicurean epistemology, as they devoted their study to the discovery of principal doctrines that could serve both, the greater understanding of the whole, and, subsequently, more specialised inquiry. We see this, too, in the works of the great natural philosophers of the Scientific Revolution—those apostates of 'Aristotelianism' and seekers of Truth whose tremendous achievement in bringing philosophy out of its darkest Age was epitomised by the efforts of Kepler and Galileo, and finally culminated in the great Newtonian synthesis; and we also find its mark in the list of brilliant minds that followed, from which the names, Darwin, Maxwell, and Einstein, stand out above all others as the most successful inferers of natural principles, who managed to deduce, from these and the greater set of physical observations, general theories of Nature that would further revolutionise our understanding of Its workings.

For along with Einstein, whose methodology we have now considered in some detail, it is equally true that Charles Darwin managed to separate himself from the common theory of his time, which required that every species of life on Earth must remain essentially unchanged subsequent to its Creation, and recognised that the awesome complexity of the great system of species could be more reasonably accounted for by a theory in which all the species have evolved from a few, or even one progenitor, according to the principle of natural selection. And, in the preface to his great *Treatise on Electricity and Magnetism*, James Clerk Maxwell laid out the method by which he formulated the electromagnetic theory, which exemplifies many aspects of the epistemological method that has now been discussed; for there he began with a statement of how he would deduce, from all the data, the general mathematical theory that would describe elecrtomagnetic phenomena with respect to observable quantities, with the aim of reducing electromagnetism to one interconnected, self-consistent dynamical theory; then, more specifically, he wrote [156],

> I have ... thought that a treatise would be useful which should have for its principal object to take up the whole subject in a methodical manner, and which should also indicate how each part of the subject is brought within the reach of methods of verification by actual measurement.
>
> The general complexion of the treatise differs considerably from that of several excellent electrical works, published, most of them, in Germany, and it may appear that scant justice is done to the speculations of several eminent electricians and mathematicians. One reason of this is that before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through Faraday's *Experimental Researches in Electricity*. I was aware that there was supposed to be a difference between Faraday's way of conceiving phenomena and that of the mathematicians, so that neither he nor they were satisfied with each other's language. I had also the conviction that this discrepancy did not arise from either party being wrong. ...

> As I proceeded with the study of Faraday, I perceived that his method of conceiving the phenomena was also a mathematical one, though not exhibited in the conventional form of mathematical symbols. I also found that these methods were capable of being expressed in the ordinary mathematical forms, and thus compared with those of the professed mathematicians.
>
> ...
>
> When I had translated what I considered to be Faraday's ideas into a mathematical form, I found that in general the results of the two methods coincided, so that the same phenomena were accounted for, and the same laws of action deduced by both methods, but that Faraday's methods resembled those in which we begin with the whole and arrive at the parts by analysis, while the ordinary mathematical methods were founded on the principle of beginning with the parts and building up the whole by synthesis.
>
> ...
>
> .... The great success which these eminent men [sc. the German electricians and mathematicians] have attained in the application of mathematics to electrical phenomena, gives, as is natural, additional weight to their theoretical speculations, so that those who, as students of electricity, turn to them as the greatest authorities in mathematical electricity, would probably imbibe, along with their mathematical methods, their physical hypotheses.
>
> These physical hypotheses, however, are entirely alien from the way of looking at things which I adopt, and one object which I have in view is that some of those who wish to study electricity may, by reading this treatise, come to see that there is another way of treating the subject, which is no less fitted to explain the phenomena, and which, though in some parts it may appear less definite, corresponds, as I think, more faithfully with our actual knowledge, both in what it affirms and in what it leaves undecided.

Thus, Maxwell recognised the philosophical brilliance of Michael Faraday, who began by considering the whole system of observed electromagnetic phenomena, and inferred a general framework for electromagnetic field theory which is consistent overall. He then applied himself to understanding the phenomena in this way, and then used mathematical expertise to analytically deduce the corresponding theory.

This is basically the same approach that has been taken more recently by David Bohm and Basil Hiley, in *The Undivided Universe*. As Hiley wrote in the Preface [180],

> I hope this book will be a fitting testimony to this very radical and original thinker [Bohm] who rejected the view of conventional quantum mechanics, not for ideological reasons, but because it did not provide a coherent overall view of nature, a feature that David felt an essential ingredient of any physical theory. It was like Escher's "The Waterfall", a fascinating picture in which region by region appeared to be carefully constructed and consistent, but when one stepped back to perceive the whole, a contradiction was there for all to see. Indeed the most radical view to emerge from our deliberations was the concept of wholeness, a notion in which a system formed a totality whose overall behaviour was richer than could be obtained from the sum of its parts.

As well, they say later on, regarding the conventional interpretation of quantum mechanics, that 'It is not that we want to dismiss the epistemological approach totally, but rather that we feel that both approaches have to be pursued and perhaps ultimately brought into relationship.' For, as they said in the introduction, one of the main advantages of their interpretation is that

> it provides an intuitive grasp of the whole process. This makes the theory much more intelligible than one that is restricted to mathematical equations and statistical rules for using these equations to determine the probable outcomes of experiments. Even though many physicists feel that making such calculations is basically what physics is all about, it is our view that the intuitive and imaginative side which makes the whole theory intelligible is as important in the long run as is the side of mathematical calculation.

One final quotation from the book should help to round off their argument, and show that it is pretty much in line with what has been written above:

> ... Heisenberg ... said very explicitly that the essential truth was in the mathematics. This view has become the common one among most of the modern theoretical physicists who now regard the equations as providing their most immediate contact with nature (the experiments only confirming or refuting the correctness of this contact). On the other hand, in the past we began talking about our concept of physical reality and used the equations to talk about them.
>
> One of the reasons that our interpretation of the quantum theory may not have been so well understood is perhaps that it is not commonly realised that we have a rather different attitude to the mathematics. Our approach is not simply a return to the notion that the mathematics merely enables us to talk about the physical concepts more precisely. Rather we feel that these two kinds of concepts represent extremes and that it is necessary to be in a process of thinking that moves between these extremes in such a way that they complement each other.
>
> We fully accept that progress can be made from the mathematical side alone, but we feel that to stick indefinitely to this approach is too limited. ... we do not regard ... physical concepts as mere imaginative displays of the meaning of our equations. Instead, as we shall see in this chapter, we can then move to the other side of this process of thought and take our physical concepts as a guide for the development of new equations. In this way we are engaged in at least the beginning of a kind of thinking in which the clue for a creative new approach may come from either side and may flow indefinitely back and forth between the two sides.

Actually, the most significant difference between Bohm and Hiley's ontological outlook and the one taken here, is that, although they defend the concept of an ultimate Reality, they argue by inductive inference that it is 'unlimited and unknown', whereas we shall take the alternate view-point, that there are well-defined, and therefore potentially definable, Truths of Nature, which could be stated as mathematical principles of natural philosophy. But although such principles are essential to any complete physical theory, regardless of how they might be connected to an ultimate Reality, it is also recognised, as a basic property of logical inquiry, that we can never be absolutely certain whether the principles or axioms we've assumed are actually True.

Perhaps the clearest systematic application of the method discussed here, which must be used in order to determine not merely a physical theory that would describe the observed phenomena, but one that works from a clear realisation of basic natural principles to deduce a complete and self-consistent theory, which is nevertheless fully aware of its weaknesses, is found in Isaac Newton's *Philosophiæ Naturalis Principia Mathematica*. Newton begins the *Principia*, following some definitions of less commonly known physical quantities, such as mass, momentum, inertia, mechanical forces, etc., and a description of what he would precisely mean by the better-known concepts of time, space, place, and motion, with a statement of three basic mathematical principles, his famous Laws of motion, before going on, in the first two books, to use them in order to describe in detail the kinematical motions of bodies. Then, he begins the third book with the statement [3],

> In the preceding books I have laid down the principles of philosophy; principles not philosophical but mathematical: such, namely, as we may build our reasonings upon in philosophical inquiries. These principles are the laws and conditions of certain motions, and powers or forces, which chiefly have respect to philosophy; but, lest they should have appeared of themselves dry and barren, I have illustrated them here and there with some philosophical scholiums, giving an account of such things as are of more general nature, and which philosophy seems chiefly to be founded on; such as the density and the resistance of bodies, spaces void of all bodies, and the motion of light and sounds. It remains that, from the same principles, I now demonstrate the frame of the System of the World. ...

He eventually moves on to write down four Rules of reasoning in philosophy:

I 'We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.'

II 'Therefore to the same natural effects we must, as far as possible, assign the same causes.'

III 'The qualities of bodies, which admit neither intensification nor remission of degrees, and which are found to belong to all bodies within reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever.'

IV 'In experimental philosophy we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypotheses that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate, or liable to exceptions.'

Then, after a lengthy analysis involving his Law of universal gravitation, he concludes with a general scholium, where he criticises René Descartes' vortex theory, and then adds some unjustified speculation which I'll have to come back to, before concluding with,

> . . . hitherto I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses; for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimenal philosophy. In this philosophy particular propositions are inferred from the phenomena, and afterwards rendered general by induction. Thus it was that the impenetrability, the mobility, and the impulsive force of bodies, and the laws of motion and of gravitation, were discovered. And to us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of the sea.
>
> And now we might add something concerning a certain most subtle spirit which pervades and lies hid in all gross bodies; by the force and action of which spirit the particles of bodies attract one another at near distances, and cohere, if contiguous; and electric bodies operate to greater distances, as well repelling as attracting the neighboring corpuscles; and light is emitted, reflected, refracted, inflected, and heats bodies; and all sensation is excited, and the members of animal bodies move at the command of the will, namely, by vibrations of this spirit, mutually propagated along the solid filaments of the nerves, from the outward organs of sense to the brain, and from the brain into the muscles. But these are things that cannot be explained in few words, nor are we furnished with that sufficiency of experiments which is required to an accurate determination and demonstration of the laws by which this electric and elastic spirit operates.

If we consider this final statement along with his Rules of philosophy, it becomes clear that Newton's theory also contained the necessary statement of a greater epistemological purpose, to continue with experimental research until such time as sufficient data existed which would indicate some direction for a subsequent reduction of principles more fundamental than the laws he was capable of inferring in his time; for he admittedly did not know the cause of those laws, and rightly framed no hypotheses to account for them. But he would have done as well to have adhered strictly to this professed stance, and not allow himself to be victimised by dogmatic prejudice into a direct contradiction; as, just prior to this concluding statement, he *had* framed a related hypothesis, which, above, I've referred to as 'unjustified speculation'—i.e., that the 'most beautiful system of the sun, planets, and comets, could only proceed from the counsel and dominion of an intelligent and powerful Being.' For, just as the law of universal gravitation, and the equivalence of gravitational and inertial mass, are better understood as they are explained by general relativity theory: so, too, within the modern theory of cosmic evolution, relativistic cosmology is found to be sufficient to give rise naturally to the great system of the *Universe*—the clusters of galaxies and all their subsystems,—as having formed dynamically since the Big Bang—which is one of the best examples of the achievement of Einstein's theory over Newton's.

In contrast to this appeal to the 'counsel and dominion' of God, in order to explain the formation of our Solar System, even though he would 'frame no hypotheses' about his law of universal gravitation, it is interesting to note the stance taken by Kepler in the introduction to his *Astronomia Nova* (where he derived the first two of his laws), in 1609 [174]:

> So much for the authority of Holy Scripture. Now as regards the opinions of the saints about these matters of nature, I answer in one word, that in theology the weight of Authority, but in philosophy the weight of Reason alone is valid. Therefore a saint was Lactantius, who denied the earth's rotundity; a saint was Augustine, who admitted the rotundity, but denied that antipodes exist. Sacred is the Holy Office of our day, which admits the smallness of the earth but denies its motion: but to me more sacred than all these is Truth, when I, with all respect for the doctors of the Church, demonstrate from philosophy that the earth is round, curcumhabited by antipodes, of a most insignificant smallness, and a swift wanderer of the stars.

Or, we might consider, from Galileo's letter of acknowledgement to Kepler, upon receiving a copy of his *Mysterium Cosmographicum* in 1597 (a rare courtesy that was paid in that direction) [174],

> ...I indeed congratulate myself on having an associate in the study of Truth who is a friend of Truth. For it is a misery that so few exist who pursue the Truth and do not pervert philosophical reason. ...

To be sure, though: Kepler's contributions to modern science were not the results of strict adherence to this dictum, from which he frequently departed [174]; whereas Newton's speculation about the formation of the Solar System was rather an exception to the rule.[8] And aside from that, it should be noted—say, according to his Rules of reasoning in philosophy, which we see no cause to go against,—that Newton, with good reason, also supported the assumption that there *is* some basic physical Truth which holds always, and that there are, therefore, basic Natural Principles which are properties of that Truth; i.e., that there are basic physical Truths which hold consistently in Reality, throughout space and time. This, we'll refer to as the principle of Natural Order. Accordingly, we define a *natural principle* as a (theoretic) principle derived through the inductive process of Heraclitean introspection, or insightful rationalisation, which might later be reduced to a principle which is logically more applicable to the scope of scientific understanding. We then deduce our theories based on the further assumption that these principles *could* be Natural Truths, and, if the resulting theory is consistent with all known natural principles, we use the scientific method to examine it for overall self-consistency, and to further broaden its scope.

However, if the theory is not consistent with all known natural principles, we recognise a significant philosophical problem, like any other we might discover through scientific investigation. We say that any theory, however powerful it may be, shall not be considered even provisionally correct unless it is consistent with every natural principle and fact within the current field of knowledge, and then only as long as new problems are not discerned; i.e.—borrowing from Leibniz' principle of sufficient reason, and from metaphysical naturalism in general, we assume an epistemological *principle of natural cause* (which is similarly expressed in Newton's Rules of philosophical reasoning),—that *there must be a basic reason for every fact of observation which is ultimately consistent with everything else we can know of Nature, and is sufficient to explain why it should be so, and not otherwise.*

[8] As one of Kepler's biographers has remarked, 'Not the least achievement of Newton was to spot the Three Laws in Kepler's writings, hidden away as they were like forget-me-nots in a tropical flowerbed' [174].

Thus, a challenge comes down to us: that, just as Maxwell did, we must abandon that inherent stubborn and naïve optimism of prejudiced thinking, that allows us to be convinced of the absolute Truth of any basic metaphysical truth of any current scientific paradigm, even in any respect we are *certain* we know it to be true, which is a constant crutch, and a general oppressor of the search for more basic knowledge,—and replace this with a rational optimism and clear acknowledgement that our theories *could* be True, as that may be warranted, but that theoretical truth must be always carefully scrutinised as our theories continue to develop; i.e., that along with making inductive interpretations of scientific observations, we must also work by continually reasoning our way towards truths which are consistent with increasingly large sets of empirical data. For even when a scientific theory is known to be entirely self-consistent, and is thus to be considered theoretically true, as it is in accord with all current observations and natural principles, we cannot ever treat it as the absolute Truth, because it will remain the product of logical analysis. This was wonderfully expressed by Einstein in the opening of his centenary commemoration of Maxwell, when he said [181],

> The belief in an external world independent of the percipient subject is the foundation of all science. But since our sense-perceptions inform us only indirectly of this external world, or Physical Reality, it is only by speculation that it can become comprehensible to us. From this it follows that our conceptions of Physical Reality can never be definitive; we must always be ready to alter them, to alter, that is, the axiomatic basis of physics, in order to take account of the facts of perception with the greatest possible logical completeness. ...

But the reassurance of such provisional epistemological completeness, as, e.g., the Newtonian theory had once possessed, as far as it was capable of accurately describing various idealised phenomena according to four basic Laws, is currently not afforded by either of the two great theories of modern physics: quantum theory, which of course does work from some principles, and which has been extremely successful, possesses no basic natural principles, and is therefore subject to multiple possible interpretations, and many paths of reduction have thus been proposed—i.e., it is not a working *metaphysical* theory;—general relativity theory, on the other hand, which was derived from natural principles, and has been equally as successful, is problematic in many ways. For these reasons, however, the methods of natural epistemology should experience far more difficulty in trying to address quantum theory directly, because we really don't have any certain indication as to its actual connection to Physical Reality, than in assessing general relativity theory, the problems of which arose because it incessantly *wants* to be something other than what Einstein tried to make it, which is basically unnatural—i.e., although it is a deductive theory, it is so *only* in the same fashion as the Eleatic theory, as it was never properly reconciled with Heraclitus' principle of flux, but holds fast to a seemingly natural principle that will likewise have to be abandoned once we've examined the requirements of Real dynamics more carefully. For as it currently stands, relativity is not a *natural* metaphysical theory, because it is basically inconsistent with the principle of natural cause.

This problem has always been known, and many attempts over the years have tried to render relativity theory more natural than it is described in Einstein's (or, rather, Minkowski's) interpretation, by those who have recognised how unnatural his theory really was made to be. Principal among these, I think, are the works of Weyl [22], Milič Čapek [182], and Stein

[9, 10], as well as the geometrodynamics theory, described in [183, 184] and the works of John Wheeler throughout the 1960s—see [185] for a comprehensive account of this theory from 1971. However, the general problem with these descriptions, I believe, is that although all fully realised the natural contradiction in Einstein's initial interpretation, none of these found a way to clearly discern what was really always at the heart of the problem—although I think Weyl, who, in [22], did examine the natural principles of the theory in the way just described, came very close, and probably had roughly the correct intuitive understanding, though it had not been sufficiently explicit.

Of course, if the different interpretation of relativity theory we are now set to describe, is to be consistent with the current paradigm—in line with what it wants to be, but has not actually achieved, as I've just said,—it must not merely be self-consistent, but consistent with the other theories that contribute to that paradigm; specifically, it would be a bad start to begin by being inconsistent with quantum theory, as relativity theory is commonly thought to be. So, as the problem currently under investigation is the Nature of spacetime, and as we currently do not have any *good* reason to believe otherwise, we shall provisionally avoid the problem of not knowing what quantum theory Really is, and therefore work towards developing a theory that may, for all we currently know, be potentially True, and thus, potentially reconcilable with a more complete understanding of quantum theory, in the following way:—we propose, as basic natural principles, the subsistence and persistence of Reality; furthermore, we will assume of that Reality, that matter and space(time) coexist in independently ontologically Real states, that space(time) is *complete*, in the formal sense of metric spaces, and that the quantum field incidentally describes material interactions *only*; i.e., that space(time) is *not* quantised, though particle interactions may be described as such. Accordingly, by considering only the implicit Nature of spacetime to begin with, we shall not find a need to directly address quantum mechanics from the outset.[9]

The consistency of these assumptions shall not be directly addressed until much later in the analysis, but they are to be considered, for now, as forming a basic working principle from which we will eventually derive a theory that can be shown to be truly consistent in so many ways with the novel facts, theoretical refinements, and *implicit* resolution of conceptual problems of modern physics. So now, keeping in mind both this principle and the greater epistemology to which it has been decided that all natural metaphysical theories must ultimately be objectively assessed, we shall return to our discussion of the historical progression of metaphysical theories concerning the Nature of Time.

2.5 On Ancient Atomism

We've seen that in his time, Parmenides had renounced Time, and thus effectively put an end to *purpose* with his logical argument, if a hole could not be found in it. For if all is One, then the will that drives our efforts is purely artificial or illusory. But although the Eleatic doctrine would maintain a healthy tradition to this day, that vital snag, as I've mentioned

[9]It is worth noting, that in their attempt to construct a quantum theory upon very similar grounds, Bohm and Hiley wrote, 'If this theory is intended to apply cosmologically, it is evidently necessary that we should not, from the outset, assume essential elements that are not capable of being included in the theory' [180].

above, was not long in coming. Not surprisingly, though, it came from Leucippus, who many accounts tell was to have come from Elea, where he had been a student of either Zeno or Melissus, who are Parmenides' two most famous followers. He is reported to have left Elea and moved to Abdera, where he developed the first atomic theory with Democritus. What the Atomists did, provides another example of the method that has just been discussed: like many others, they realised the absurdity of the Eleatic theory, as it had provided no intuitive argument to refute Heraclitus' principle; therefore, they attacked its basic principle, saying that because change and motion surely must be Real, what-is-not *is*, and that what-is moves through it, changing as it does so—i.e., that reality, in their theory, consisted of physical corporeal matter moving about in a physical incorporeal void.

Although none of their original works now remain, the theory they developed, as well as their mode of reasoning, were well-documented through the works of Aristotle; so we are able to present their theory through such quotations. Continuing, then, in line with this brief introduction to the theory, we have, from Simplicius of Cilicia's sixth century A.D. commentary on Aristotle's *Physics*, 28.4-27 [166],

> Leucippus of Elea or Miletus (both are reported) associated in philosophy with Parmenides, but did not follow the same path as Parmenides and Xenophanes about what there is, but, it seems, the opposite. For they made the universe consist of one changeless, ungenerated, and bounded thing and did not admit that one could even think of what is not, while he posited the atoms, an infinite number of elements in continual motion, and held that they have an infinite number of shapes, since there is no more reason for them to be one shape than another, and that coming to be and change are unceasing among the things that there are. Further, that which is exists no more than that which is not, and both are alike causes of the things that come to be. For positing the nature of the atoms as a solid and a plenum he said that it is what is and that it travels about in the void, which he called 'what is not' and said that it is no less than what is.
>
> Similarly his associate Democritus of Abdera posited as principles the plenum and the void, calling the one 'what is' and the other 'what is not.' They posit the atoms as matter of the things that there are, and generate everything else by their differentiating characteristics. There are three of these, 'rhythm,' 'turning,' and 'contact,' which is to say shape, position and arrangement. For it is natural for like to be affected by like and for things of the same kind to move towards one another and for each of the shapes to be reorganized into a different complex and so make another state. So they claimed that, the principles being infinite, they would readily account for all substances and properties, and what is the cause of anything and how. That is why they say that only those who make their elements infinite can make everything turn out according to their theory. They say that the number of shapes belonging to the atoms is infinite because there is no more reason for them to be one shape than another; for that is the explanation they give of the infinity.

Then, from his commentary on Aristotle's *On the Heavens*, 294.33-295.36 ([4], *F 208* R^3), we have the further description,

> A few words quoted from Aristotle's *On Democritus* will reveal the line of thought of those men [sc. the Atomists]:—Democritus thinks the nature of the eternal entities consists of small substances infinite in number; he supposes a place for them, different from them and infinite in extent, and to this he applies the names 'void', 'nothing', and 'the infinite', while to each of the substances he applies the names 'thing', 'solid', and 'existent'. He thinks the substances are so small as to escape our senses, but have all sorts of shapes and figures, and differences of size. From these, then, as from elements, are generated and compounded visible and perceptible masses. The substances are at variance and move in the void because of their dissimilarity as well as the other aforesaid differences, and as they move they collide with each other and interlock in such a way that, while they touch and get close to each other, yet a single substance is never in reality produced from them; for it would be very simple-minded to suppose that two or more things could ever become one. The cause of these substances remaining with one another for some time he ascribes to the bodies fitting into one another and catching hold of one another; for some of them are scalene, others hook-shaped, others concave, others convex, and others have countless other differences. He thinks that they cling to one another and remain together until some stronger necessity arriving from the environment scatters them apart and separates them. He ascribes the genesis and separation opposed to it not only to animals but also to plants, and to worlds, and generally to all perceptible bodies.

Finally, we'll consider one of the more elaborate descriptions of the Atomists' theory given by Aristotle himself, which compares it to other contemporary theories, and, incidentally, indicates that the concept of a void was not so much a subsequent realisation of theirs, thereafter understood as a sufficient resolution of the problem posed by the Eleatic theory, as it was actually commonly considered a significant problem throughout Ancient natural philosophy, at least from the time of the Eleatics, and which very much remained a contentious point in Aristotle's time, over a hundred years later—though, probably as a direct result of the Eleatics' reasoning, the Atomists *were* the first to have taken the definitive stance directly opposed to theirs, and to have deduced a theory that was equally consistent with its principles. To be sure: although modern physics provides overwhelming confirmation that vacuous regions of spacetime do actually exist, it is still not a universal belief that spacetime should subsist independent of any material it might contain; so it is, that, according to the principles stated above, we have here taken the same stance that was first truly supported in ancient atomism:—that the incorporeal void of spacetime is to be a genuine component of reality. The relevant quotation, from Aristotle's *On Generation and Corruption*, runs thus [4] (324b25-325b15):

> Some philosophers think that the last agent—the agent in the strictest sense—enters in through certain pores, and so the patient suffers action. It is in this way, they assert, that we see and hear and exercise all our other senses. Moreover, according to them, things are seen through air and water and other transparent bodies, because such bodies possess pores, invisible indeed owing to their minuteness, but close-set and arranged in rows—the more transparent the body, the

more so.

Such was the theory which some philosophers (including Empedocles) advanced in regard to the structure of certain bodies. They do not restrict it to the bodies which act and suffer action; but combination too, they say, takes place only between bodies whose pores are in reciprocal symmetry. The most systematic theory, however, and one that applied to all bodies, was advanced by Leucippus and Democritus: and, in maintaining it, they took as their starting-point what naturally comes first.

For some of the older philosophers[10] thought that what is must of necessity be one and immovable. The void, they argue, is not; but unless there is a void with a separate being of its own, what is cannot be moved—nor again can it be many, since there is nothing to keep things apart. And they hold that the view that the universe is not continuous but consists of separate things in contact is no different from the view that there are many (and not one) and a void. For if it is divisible through and through, there is no one, and no many either, but the Whole is void; while to maintain that it is divisible at some points, but not at others, looks like an arbitrary fiction. For up to what limit is it divisible? And for what reason is part of the Whole indivisible, i.e. a *plenum*, and part divided? Further, they maintain, it is equally necessary to deny the existence of motion.

Arguing in this way, therefore, they were led to transcend sense-perception, and to disregard it on the ground that one ought to follow reason; and so they assert that the universe is one and immovable. Some of them add that it is infinite, since the limit (if it had one) would be a limit against the void.[11]

There were, then, certain thinkers who, for the reasons we have stated, enunciated views of this kind about the truth ... Moreover,[*] although these opinions appear to follow logically, yet to believe them seems next door to madness when one considers the facts. For indeed no lunatic seems to be so far out of his senses as to suppose that fire and ice are one: it is only between what *is* right, and what *seems* right from habit, that some people are mad enough to see no difference.

Leucippus, however, thought he had a theory which harmonized with sense-perception and would not abolish either coming-to-be and passing-away or motion and the multiplicity of things. Making these concessions to the phenomena and conceding to the Monists that there could be no motion without a void, he states that void is not-being, and no part of what is is not-being; for what is in the strict sense of the term is an absolute *plenum*. This *plenum*, however, is not one: on the contrary, it is a many infinite in number and invisible owing to the minuteness of their bulk. The many move in the void (for there is a void); and by coming together they produce coming-to-be, while by separating they produce passing-away. Moreover, they act and suffer action wherever they chance to be in contact (for they are not thereby one), and they generate by being put together

[10] This paragraph and the next outline the argument of the Eleatics.

[11] This was a contribution of Melissus'; see, e.g., [165] for further description.

[*] 'One or more arguments against the Eleatic theory appear to have dropped out' (Joachim) [i.e., translator's note].

and becoming intertwined. From the genuinely one, on the other hand, there never could have come-to-be a multiplicity, nor from the genuinely many a one: that is impossible. But just as Empedocles and some of the other philosophers say that things suffer action through their pores, so all alteration and all passion take place in this way, breaking-up (i.e. passing-away) being affected by means of the void, and so too growth—solids creeping in to fill the void places.

Empedocles too is practically bound to adopt the same theory as Leucippus. For he must say that there are certain solids which, however, are indivisible—unless there are continuous pores all through the body. But this is impossible; for *then* there will be nothing solid beside the pores but all of it will be void. It is necessary, therefore, for his contiguous things to be indivisible, while the intervals between them—which he calls pores—must be void. But this is precisely Leucippus's theory of action and passion.

Such, approximately, are the accounts of the manner in which some things act while others suffer action. And as regards the Atomists, it is not only clear what their explanation is: it is also obvious that it stands in tolerable consistency with the assumptions they employ.

He then moves on to consider the inconsistency in Empedocles' theory, before coming back to compare the atomic theory to that of Plato, upon which he concludes, 'Thus the comings-to-be and the dissociations result from the indivisibles *according to Leucippus* through the void and through contact (for it is at the point of contact that each of the composite bodies is divisible), but *according to Plato* in virtue of contact alone, since he denies there is a void.'

Therefore, according to the atomic theory, the void would allow a place for every world (according to similar accounts, the theory held that there should have been infinitely many of them in the void, continually coming-to-be and passing-away) to have been in the past, and to move to in the future, while the atoms' continued motion within each world would allow for the change, locomotion, alteration, etc., of things, as we observe. The atoms were required to have been moving through the void since eternity, in order to naturally produce coming-to-be and passing-away, of everything from whole worlds to the smaller bodies within worlds; but this seems to have come at some price, as it suggests that no atom, or aggregate body, should have the ability to actually choose its path within the world, because any change that would thus occur must come about by necessity, and would be as strictly determined as in the Eleatic theory; for so the corporeal past-and-future of the Eleatic theory, having been replaced by a real void, and therefore not physically real itself, was nevertheless only changed to be an equally absolute metaphysical conception.

However, it is uncertain whether the Atomists did accept such a requirement of strict determinism in their theory, or whether they had in fact allowed the possibility that animate beings, once formed, would be capable of exercising free will, choosing their paths through the void according to reason, as much as by necessity—and arguments to support either side can easily be made from the existing fragments and testimonia; see, e.g., [166] for further discussion on this point, which I'll leave off with after two further quotations: first, Diogenes Laërtius relates Democritus' opinions as having been that [167]

> ...the atoms are unlimited in size and number, and they are borne along in the whole universe in a vortex, and thereby generate all composite things—fire,

> water, air, earth; for even these are conglomerations of given atoms. And it is because of their solidity that these atoms are impassive and unalterable. The sun and the moon have been composed of such smooth and spherical masses [*i.e.* atoms], and so also the soul, which is identical with reason. We see by virtue of the impact of images upon our eyes.
>
> All things happen by virtue of necessity, the vortex being the cause of the creation of all things, and this he calls necessity. The end of action is tranquillity, which is not identical with pleasure, as some by a false interpretation have understood, but a state in which the soul continues calm and strong, undisturbed by any fear or superstition or any other emotion. This he calls well-being and many other names. The qualities of things exist merely by convention; in nature there is nothing but atoms and void space.

And secondly, we have the one fragment that is likely attributable to Leucippus: 'Nothing happens in vain, but everything from reason and by necessity' [166]. The fact that reason, though its existence is by virtue of necessity, as with all else, was nevertheless included in this statement, suggests that it was not stated here redundantly, but actually as a basic cause of happenings.

Compared with Aristotle's later theory of 'action' and 'passion' to explain interaction within a continuous distribution of matter, i.e. with no void place anywhere, the original atomic theory, in which individual atoms would be jostled about within an incorporeal void, is far closer to the modern physical conception of reality, as it maintained that things would move about and be seen to do so as the elements of observation would physically traverse space, from one separate body to another. But consider that even to us, who, e.g., accept the concept of molecules of a gas in thermal motion as something natural, this idea in fact remains somewhat foreign; for although we are generally content to think of photons, neutrinos, and other particles travelling through practically empty space, e.g. *en route* from the Sun to the Earth, we still marvel at the fact that trillions of neutrinos generated by the Sun pass through our bodies every second—indeed, that a proportionally greater number pass through the Earth every second, which impedes them little more than if it were a pure vacuum—*only because they are electrically neutral*, and can therefore only interact with other particles when they pass at inconceivably small distances. To a neutrino, even the Earth is, in effect, a sparse collection of particles separated by huge expanses of void space.

But consider that even Einstein, through Mach's influence, had kept as close to the Eleatic view as he was able; e.g., in 'Relativity and the Problem of Space', he argued [5],

> .... Mach, in the nineteenth century, was the only one who thought seriously of an elimination of the concept of space, in that he sought to replace it by the notion of the totality of the instantaneous distances between all material points. ...
>
> [In Newtonian mechanics], "physical reality", thought of as being independent of the subjects experiencing it, was conceived as consisting, at least in principle, of space and time on one hand, and of permanently existing material points, moving with respect to space and time, on the other. The idea of the independent existence of space and time can be expressed drastically in this way: If matter

were to disappear, space and time alone would remain behind (as a kind of stage for physical happening).

...

We are now in a position to see how far the transition to the general theory of relativity modifies the concept of space. In accordance with classical mechanics and according to the special theory of relativity, space (space-time) has an existence independent of matter or field. In order to be able to describe at all that which fills up space and is dependent on the coordinates, space-time or the inertial system with its metrical properties must be thought of at once as existing, for otherwise the description of "that which fills up space" would have no meaning.* On the basis of the general theory of relativity, on the other hand, space as opposed to "what fills space", which is dependent on the co-ordinates, has no separate existence. Thus a pure gravitational field might have been described in terms of the g_{ik} (as functions of the coordinates), by solution of the gravitational equations. If we imagine the gravitational field, i.e. the functions g_{ik}, to be removed, there does not remain a space of the type [Eq. (2.1)], but absolutely *nothing*, and also no "topological space". For the functions g_{ik} describe not only the field, but at the same time also the topological and metrical structural properties of the manifold. A space of the type [Eq. (2.1)], judged from the standpoint of the general theory of relativity, is not a space without field, but a special case of the g_{ik} field, for which—for the coordinate system used, which itself has no objective significance—the functions g_{ik} have values that do not depend on the co-ordinates. There is no such thing as an empty space, i.e. a space without field. Space-time does not claim existence on its own, but only as a structural quality of the field.

When we see that even one of the greatest minds of the twentieth century, with all the knowledge of modern physics at his disposal, had propounded a block universe description in which space itself, although it could not be a plenum, should only exist as an accidental property of the more complex, scientifically advanced, four-dimensional field, it should not be difficult to see why, in the fifth century B.C., aside from the fact that the principles of Atomism gave rise to a comparably more reasonable description of perceptions of Nature, the Atomists' void would have seemed to many to be a gratuitous hypothesis, or, as Einstein still thought of it in his day, an 'absurd notion', when there would have been no obvious indication that any such thing as a Void should have independent existence in the world, seen to consist of the Earth, its atmosphere, and all the heavenly bodies.

If we examine the situation from this perspective, making a further note of consistency, that the Eleatics would have had good reason for believing in the perfect continuity of all matter, which they believed to such an extent that they accepted the consequence that motion and change would have to be denied simply because such things couldn't be Real

*If we consider that which fills space (e.g. the field) to be removed, there still remains the metric space in accordance with

$$ds^2 = {dx_1}^2 + {dx_2}^2 + {dx_3}^2 - {dx_4}^2, \tag{2.1}$$

which would also determine the inertial behaviour of a test body introduced into it.

in the infinite bottleneck, then it seems reasonable that the later abstract metaphysicians, in utilising the tools of their trade, would have seen no good reason that there should have been a void, as, in a very different way, they could sufficiently explain all of the phenomena. For then, even if matter did exist in a real physical place, it could be that this place would be continuously filled with matter, and that change occurred by rearrangement of matter in time, as witnessed by this continually changing arrangement of things. In his *Physics*, Aristotle wrote [4] (208b1-8):

> The existence of place is held to be obvious from the fact of mutual replacement. Where water now is, there in turn, when the water has gone out as from a vessel, air is present; and at another time another body occupies this same place. The place is thought to be different from all the bodies which come to be in it and replace one another. What now contains air formerly contained water, so that clearly the place or space into which and out of which they passed was something different from both.

Accordingly, e.g., even when a dry sponge is compressed, it would not be compressed into place devoid of matter, but air would have filled parts of its interior, which would then be expressed into the surroundings. Such are the reasons why Aristotle could, one hundred years later, following the Athenian progress in metaphysics, refute the existence of void, as empty place, and propound a theory in which place would be everywhere full, but still physically changing in time. Of course, his theory had its own particular set of problems, such as a requirement to account for the sensation of things that happen at a distance, which he resolved with his theory of action and passion,—as he did with each of the other problems that his theory encountered, again with tolerable consistency, even though that theory is now understood to have been completely false.

2.6 On Time

It is now common knowledge that the corporeal existence supposed in Aristotle's physical theory of Nature was incorrect; and moreover, that the Atomists had been basically on the right track. As I read the accounts of their theory, it seems to me that it is primarily lacking in the way it treats time when accounting for the processes that must take place *through* time. According to certain accounts—e.g., the cosmogony of Leucippus that was summarised by Diogenes Laërtius [167]—it should be understood that the theory described an infinite three-dimensional void containing infinitely many bounded 'worlds', of which our observable Universe would be one. However, according to the modern understanding of the large-scale Universe, due to the past century's cosmological observations, the realistic sense that It must be unbounded suggests that the Atomists' theory should have been more accurate if time had instead been described by a real fourth dimension of the void, through which each 'world', unbounded in three dimensions, would move uniformly, always remaining a smooth hypersurface. Furthermore, regarding the consistent physical progression of time, as it must take place in the three-dimensional atomic cosmology, which is formally no different a problem than it was in his own theory, Aristotle recognised that [4] (252a32-b6)

> . . . it is a wrong assumption to suppose universally that we have an adequate first principle in virtue of the fact that something always is so or always happens so. Thus Democritus reduces the causes that explain nature to the fact that things happened in the past in the same way as they happen now; but he does not think fit to seek for a principle to explain this 'always': so, while his theory is right in so far as it is applied to certain individual cases, he is wrong in making it of universal application. Thus, a triangle always has its angles equal to two right angles, but there is nevertheless an ulterior cause of the eternity, whereas principles are external and have no ulterior cause. Let this conclude what we have to say in support of our contention that there never was a time when there was not motion, and never will be a time when there will not be motion.

Aristotle's argument for a physical time in which the type of motion he means here—an abstract metaphysical motion that is (to briefly paraphrase the principal conclusions he makes in the final two books of his *Physics* [4], without including any of the logical argument) numerically one and the same, and simple; that is primary and the measure of all other motions: the continuous rotary motion imparted by a prime mover who alone remains eternally unmoved—is really not relevant for us because he denied the subsistence of the void in principle, having concluded an argument against it sometime earlier on. Instead, we will eventually come to a different argument for physical time that is more relevant to our purposes. For now, it should be sufficient to note that time must be physical, and consider some consequences that this has for the 'four-dimensional atomic theory' just described.

It is true that one problem a theory of four real dimensions would have faced, which is possibly also a reason why it was not taken up by the Atomists, is one that arises when extra dimensions need to be added in the derivation of a physical description of Nature, which was recognised by Zeno of Elea, who wrote a collection of paradoxes intended to pose difficulties for any theory that might come against the Eleatic one; the particular one of relevance here, is [4] (209a25), 'if everything that exists has a place, place too will have a place, and so on *ad infinitum*.'

However, if the three dimensions of space should be unbounded, and if time should be recognisable as the relative positioning of all matter as the Universe really moves through a physical time-dimension, due to a prior physical property of reality according to which time should continually pass with consistent measure, then this problem can be addressed by considering the only two possibilities that then exist: either time should be the result of a more abstract metaphysical cause, as both the Atomists and Aristotle appear to have thought it was, witnessed through the continual rearrangement of the Universe (whether or not there would be physical void separating atoms in space), which still needs to be addressed in the physical theory, as Aristotle seems to be the first to have recognised; or, time is an extra dimension of a physical void, through which every body in the Universe moves continuously, as they also may rearrange themselves in space (which, again, need not be full). But these two cases are obviously not formally distinct, as they both describe time as the motion of a universe through something physical and incorporeal, and the choice, whether to conceive of that as abstract or not, is purely arbitrary.

However, there is another, accidental sense of time in this one formal description: the non-real (purely *ideal*) one that the Atomists probably had in mind, which describes the past, present, and future events that happen in the Universe, as they actually were, are, and will be. With the addition of this dimension, the physical description of all events that would occur as a three-dimensional universe really moved through a real four-dimensional void, would not be four-dimensional (as they would be if that real fourth dimension were neglected, and only the events occurring in the universe through time were taken into account), but actually five-dimensional; however, because the Universe is still supposed to be three-dimensional, the set of events that occur in the Universe itself, as it moves through that fourth real dimension, would only be a four-dimensional slice of that five-dimensional description of everything that ever was and will be.

Actually, the fact that this ideal time dimension exists in the description of any such theory—i.e., even in Aristotle's theory, in which time is the uniform motion imparted on everything by a prime mover, which takes place through an abstract metaphysical dimension—implies a second option, which was left out above by requiring *real* movement through the fourth physical dimension; viz., that real motion does not actually need to take place, but could really be that same accidental dimension in which all past and future events in the Universe occur—the physical description of all occurrences in the Universe—which is simply promoted in status, to be the actual fourth dimension of physical reality—as it was in the Eleatic theory, and as it is described in the original interpretation of relativity theory: a physical four-dimensional block.

So we have two formally distinct descriptions of Time, according to the reasonings of the ancient Greek philosophers. However, as shown in the detailed discussion in §§ 3.1 – 3.2, relativity theory cannot possibly distinguish between these two interpretations of physical reality, because the four-dimensional continuum of events that the formal theory is used to describe—viz., *spacetime*—can always be thought of, either as a block, like the Eleatic One, or as the set of events occurring on a well-defined three-dimensional space dynamically moving through a four-dimensional void, according to Einstein's equation, as in our 'four-dimensional atomic theory'; then, it is an immediate consequence that in the latter case the fifth (physical) dimension is not required to be *real*, as the regular evolution of all four dimensions of reality must progress according to the formal dynamics that need to be assumed in the physical theory, and Zeno's infinite sequence of real places is no more required than it is in the Eleatic theory.

Now, it is important to investigate in more detail the physical reasons why Time actually needs to be a physical dimension, as It is formally described by the theories of, e.g., Aristotle, Newton, and Einstein; viz., as the critical philosophical argument that Time must be somehow physical—i.e., subject to the fundamental Laws of physics,—originates in the fact, common to both the pre-relativistic and relativistic time concepts, that Its passage is always *perceived* with consistent measure. A sufficient argument for this was already given—though he did not draw the same natural conclusion from his reasonings as we shall—by Augustine of Hippo, near the end of the fourth century A.D., in the eleventh Book of his *Confessions*, where he wrote of his personal reflections on the Nature of Time, which Russell described as 'a very admirable relativistic theory of time' [186]. Augustine begins with a speculation on the nature of God's own existence, which Russell paraphrases [186]:—

> Why was the world not created sooner? Because there was not "sooner." Time was created when the world was created. God is eternal, in the sense of

> being timeless; in God there is no before and after, but only an eternal present. God's eternity is exempt from the relation of time; all time is present to Him at once. He did not *precede* His own creation of time, for that would imply that he was in time, whereas He stands eternally outside the stream of time.

This, Russell points out, is what leads Augustine to his 'admirable relativistic theory':

> "What, then, is time?" he asks. "If no one asks of me, I know; if I wish to explain to him who asks, I know not." Various difficulties perplex him. Neither past nor future, he says, but only the present, really *is*; the present is only a moment, and time can only be measured while it is passing. Nevertheless, there really is time past and future. We seem to be led into contradictions. The only way Augustine can find to avoid these contradictions is to say that past and future can only be thought of as present: "past" must be identified with memory, and "future" with expectation, memory and expectation being both present facts. There are, he says, three times: "a present of things past, a present of things present, and a present of things future." "The present of things past is memory; the present of things present is sight; and the present of things future is expectation." To say that there are three times, past, present, and future is a loose way of speaking.
>
> He realizes that he has not really solved all difficulties by this theory. . . . But the gist of the solution he suggests is that time is subjective: time is in the human mind, which expects, considers, and remembers. It follows that there can be no time without a created being, and that to speak of time before the Creation is meaningless.
>
> I do not myself agree with this theory, in so far as it makes time something mental. But it is clearly a very able theory, deserving to be seriously considered. I should go further and say that it is a great advance on anything to be found on the subject in Greek philosophy. It contains a better and clearer statement than Kant's of the subjective theory of time—a theory which, since Kant, has been widely accepted among philosophers.

Russell goes on to conclude that, since Augustine's theory anticipated not only Kant's theory of time, but Descartes' *cogito*, he deserves a high place as a philosopher. But what Russell fails to relate here, which must be of far greater interest to the physicist, is the importance of the evidences Augustine presents, which lead him to his conclusion. For everyone knows that any conclusion reasoned inductively is not necessarily the only conclusion that *can* be drawn from such a process; and, as Maxwell wrote, 'It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in the nascent state' [156]. In this case, it should actually become clear to any relativist, that in his loose way of speaking about past, present, and future, as they relate to the measurements of temporal durations by a single observer, Augustine had anticipated modern physics in providing a good basis for certain conceptions of *spacetime* and *proper time*; therefore, his theory is far more relativistic than Russell leads us to believe, as it appears his comment was probably meant to address, instead, the less interesting fact that, generally speaking, the eternal realm of God through which our minds progress, is essentially the reality to which Einstein had attributed the description given by relativity theory.

The passages of Augustine's text relating these progressive ideas, are [187]:

> . . . we take such measure of the times in their passing by, as we may be able to say, this time is twice so much as that one; or, this is just so much as that: and so of any other parts of time, which be measurable. We do therefore, as I said, take measure of the times as they are passing by. And if any man should now ask me: How knowest thou? I might answer, I do know, because we do measure them: for we cannot measure things that are not; and verily, things past and to come are not. But for the present time now, how do we measure that, seeing it hath no space? We measure it therefore, even whilst it passeth, but when it is past, then we measure it not: for there will be nothing to be measured. But from what place, and by which way, and whitherto passes this time while it is a measuring? Whence, but from the future? Which way, but through the present? Whither, but into the past? From that therefore, which is not yet: by that, which hath no space: into that, which is not still. Yet what is it we measure, if not time in some space? For we use not to say, single, and double, and triple, and equal, or any other way we speak of time, but with references still to the spaces of times. In what space therefore do we measure the time present? Whether in the future space, whence it is passing? Or in the present, by which it is passing? But no-space we do not measure. Or in the past, to which it passeth? But neither do we measure that which is not still.
>
> . . .
>
> . . . . I will not therefore demand now what that should be which is called day: but, what time should be, by which we measuring the circuit of the sun, should say, that he had then finished it in half the time he was wont to do, if so be he had gone it over in so small a space as twelve hours come to: and when upon comparing of both times together, we should say, that this is but a single time, and that a double time, notwithstanding that the sun should run his round from east to east sometimes in that single time, and sometimes in that double time. . . . I perceive time therefore to be a certain stretching. But do I perceive it, or do I seem to perceive it? . . .
>
> . . . . Seeing therefore the motion of a body is one thing, and that by which we measure how long it is, another thing; who cannot now judge which of the two is rather to be called time? For and if a body be sometimes moved uncertainly, and stands still other sometimes; then do we measure, not his motions only, but his standing still too: and we say, It stood still as much as it moved; or It stood still twice or thrice so long as it moved; or any other space which our measuring hath either perfectly taken, or guessed at, more or less, as we use to say. Time therefore is not the motion of a body.
>
> . . .
>
> . . . it seems to me, that time is nothing else but a stretching out in length; but of what, I know not, and I marvel, if it be not of the very mind. For what is it, I beseech thee, O my God, that I now measure, whereas I say, either at large, that this is a longer time than that: or, more particularly, that this is double to that? I know it to be time that I measure: and yet do I neither measure the time

> to come, for that is not yet: nor time present, because that is not stretched out in any space: nor time past, because that is not still. Is it the times as they are passing, not as they are past? For so I was saying.

From these speculations, the progression of Augustine's thoughts continue towards that which must have seemed to him to be the simplest conclusion, and therefore, according to the need for logical economy, the most acceptable one: viz., that the 'space of time' might be mental. We are quite sure now, however, that this is not the case,—that time itself is also a real physical dimension of a greater subsistent field; but it is also obvious that a physical reality of time is the natural conclusion of Augustine's speculations, regardless of whether that would be only mental.

On the other hand, his argument in no way implies that he has thought of any difference in observation of the passages of time measured by two relatively moving observers; nor that time and space should form something we could recognise as a topologically connected, four-dimensional space; thus, Augustine's reasoning applies sufficiently well to the concept in Newton's theory, that [3]

> Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration: relative, apparent, and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a day, a month, a year.

This should, however, not be thought of as a mark against the theory, but merely as an incompleteness, as his idea is in fact more basic—merely that there is, all ways, a prior, consistently measurable duration—which is why it pertains indiscriminately to both the pre-relativistic and relativistic time concepts; i.e., we can say that time, according to Augustine's theory, is something that *is* extended, in whatever capacity that may be; and that it is measurable by any observer, who, for themself, must be sensible of a dynamic present, as well as its associated past and future, in which durations of processes within proper three-dimensional space can be observed, but which exists prior to any measurable occurence, underlying the very idea of process—whether that would be observed as active or inertial.

Thus, by the very nature of his argument, in which he considers time only from the subjective standpoint of personal observation, and does not consider how its measurement would differ according to the perspective of an external body—e.g., which might move some distance, then sit awhile, then perhaps spin in place,—regardless of the fact that he did not realise the additional complication of relative motion, Augustine did find a good description of time, witnessed from the proper perspective as some basic physically extended aspect of Nature, which may or may not be absolute, and thus could, as well as not, be used to form a basis of an underlying four-dimensional, connected spacetime. Augustine's contribution to the natural philosophy of time can therefore be stated: He solidified Heraclitus' principle of uniform flux according to a further rationalisation of common sense which led him to realise that this natural progression through time always appears to have consistent measure, and must therefore be physical, laid out along a physical dimension.

So we see the relevant issue; viz., that Augustine's theory, though generally applicable to more developed physical theories, actually did not go far enough to be considered a completed theory of time. And, of course, we know that the final essential advancement that was made on this singular perspective came only fifteen hundred years later, after the relativity of inertia had emerged from Renaissance thought, and, furthermore, following the development of a theory of light which culminated in the precision measurement, against conventional reason, that its speed should be the same constant and finite value in all frames,—when Einstein, in postulating natural principles relating to the kinematics of these observations, realised the physical consequences pertaining to measures of time and simultaneity, when two relatively moving observers' prespectives would then be considered in relation to one another; viz., that the two observers would not agree on the simultaneity of events perceived in their proper frames.

2.7 The Original Interpretation of Spacetime

The relativity of simultaneity requires that the proper coordinate system of every relatively moving frame must provide a relatively unique description of reality; and the differentiating factor between the interpretation of relativity theory to be discussed in this section, and the new one that is described in Chapter 3, is in fact how that reality should be formally identified in the mathematical description, if it should actually be dynamic.

The original interpretation of spacetime, which has remained a central aspect of the theory, was first described in 1908, by Einstein's former mathematics teacher from his days in Zürich, Hermann Minkowski, who, at a conference in Cologne, proposed both: that relativity theory should be taken to mean that space and time are not disjoint, but actually form a four-dimensional continuum; and, subsequently—in a metaphysical statement equvalent to one for which McTaggart argued at length in 'The Unreality of Time', published the following month [188],—that this should be the True aspect of Reality, in which all events exist immediately, throughout the entire continuum, once and for all. In his words [6],

> ...space by itself, and time by itself, are [henceforth] doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.
>
> ...
>
> .... A point of space at a point of time, that is, a system of values x, y, z, t, I will call a *world-point*. The multiplicity of all thinkable x, y, z, t systems of values we will christen the *world* .... Not to leave a yawning void anywhere, we will imagine that everywhere and everywhen there is something perceptible .... Then we obtain, as an image, so to speak, of the everlasting career of the substantial point, a curve in the world, a *world-line*, the points of which can be referred unequivocally to the parameter t from $-\infty$ to $+\infty$. ...
>
> ...
>
> .... We should then have in the world no longer *space*, but an infinite number of spaces, analogously as there are in three-dimensional space an infinite number of planes. Three-dimensional geometry becomes a chapter in four-dimensional physics. Now you know why I said at the outset that space and time are to fade away into the shadows, and only a world in itself will subsist.
>
> ...

> Lorentz called the t' combination of x and t the local time of the electron in uniform motion, and applied a physical construction of this concept, for the better understanding of the hypothesis of contraction. But the credit of first recognizing clearly that the time of the one electron is just as good as that of the other, that is to say, that t and t' are to be treated identically, belongs to A. Einstein [189]. Thus time, as a concept unequivocally determined by phenomena, was first deposed from its high seat. Neither Einstein nor Lorentz made any attack on the concept of space, perhaps because in the ... special transformation, where the plane of x', t' coincided with the plane of x, t, *an interpretation is possible* by saying that the $x-$axis of space maintains its position. [Emphasis added.]

Thus, in the tradition of the Eleatics, Minkowski placed the mathematical description of any physical *process*, whether 'active' or 'inertial', in the same realm that St. Augustine had placed God. As he appears to have realised roughly the additional possibility of the interpretation that will be argued for in Chapter 3, we might wonder whether his reasons for decrying this interpretation should have had something to do with the simple facts, that the interpretation he chose to propound is mathematically more neat, or that he—again, in the Eleatic tradition—clearly abhorred the possibility of a reality in which there could physically be a 'yawning void', and so impressed that bias on natural science. Regardless, the interpretation he did decide on is the one that stuck—at least, it did with Einstein.

For it is well-documented that although Einstein did not immediately take to Minkowski's interpretation, he did come around to accepting it, as he realised the necessity of four-dimensional spacetime in order to arrive at the general theory. Pais [58] notes that Einstein once remarked to V. Bargmann that he 'regarded the transcriptions of his theory into tensor form as "überflüssige Gelehrsamkeit," (superfluous [erudition]).' It seems likely that this sentiment comes from one that Einstein is known to have carried with him from his youth; e.g., in *Albert Einstein: A Documentary Biography* [190], Carl Seelig provides a quotation from Einstein's autobiography, which seems to me to shed some more light on the comment made to Bargmann:

> ... my intuition in the mathematical field was not strong enough to be able to distinguish with basic conviction the fundamentally important from the rest of the more or less dispensable erudition. Moreover my interest in acquiring a knowledge of Nature was infinitely stronger and as a student it was not clear to me that the approach to a deeper knowledge of the principles of physics was bound up with the most intricate mathematical methods. This only dawned on me after years of independent scientific work.
>
> ... In [physics], however, I soon learned to sense those [data] in particular that could lead to fundamentals and to overlook a host of others that filled the mind and distracted from the essential.

Seelig then goes on to describe Minkowski's interpretation, noting that, in 1916, Einstein had paid tribute to its contribution to the development of general relativity theory; after this, however, he notes that 'Einstein once joked about the mathematical finesse with which his work was investigated and expanded. He said with a sigh: "Since the mathematicians have attacked the relativity theory, I myself no longer understand it any more."'[12]

This is all closely related to something Philipp Frank had written five years earlier, in *Einstein: His Life and Times* [192]:

> Einstein was at first skeptical about the use of very advanced mathematics in developing physical theories. When in 1908 Minkowski showed that Einstein's special theory of relativity could be formulated very simply in the language of four-dimensional geometry, Einstein had regarded this as the introduction of an involved formalism by which it became rather more difficult to grasp the physical content of the theory. When Max von Laue, in the first comprehensive book on Einstein's relativity theory, presented it in a very elegant mathematical form, Einstein remarked at that time jokingly: "I myself can hardly understand Laue's book."
>
> The center of German mathematical teaching and research during this period was the University of Göttingen. Minkowski taught there, and the mathematical formulation of the relativity theory had begun there. Einstein once remarked playfully: "The people in Göttingen sometimes strike me, not as if they wanted to help one formulate something more clearly, but instead as if they wanted only to show us physicists how much brighter they are than we." Nevertheless, the greatest mathematician in Göttingen, David Hilbert, realized that while Einstein did not care for the superfluous formal difficulties in mathematics, he did know how to use mathematics where it was indicated. Hilbert once said: "Every boy in the streets of our mathematical Göttingen understands more about four-dimensional geometry than Einstein. Yet, despite that, Einstein did the work and not the mathematicians."

It is clear that Einstein did not immediately intuitively accept the mathematical formalism that he would eventually use to develop the general theory; but regardless of how he came to accept it, we know that he did eventually entirely support this interpretation of physical reality. For example, when he wrote of the innovation of relativity theory in *The Meaning of Relativity*, which was first published in 1921, he said [89],

> In the pre-relativity physics space and time were separate entities .... One spoke of points of space, as of instants of time, as if they were absolute realities. It was not observed that the true element of the space-time specification was the event specified by the four numbers x_1, x_2, x_3, t. The conception of something happening was always that of a four-dimensional continuum; but the recognition

[12]This redundancy makes the English translation of Einstein's remark seem strikingly poor. Furthermore, there is total ambiguity as to whether Seelig heard this directly from Einstein, or whether this information had come from some other source, which he neglected to reference. Therefore, it is important to refer back to the original german text, in *Albert Einstein und die Schweiz* [191], which reads, 'Scherzhaft soll Einstein einmal angesichts der mathematischen Finessen, mit dennen seine Arbeiten untersucht und erweitert wurden, geseufzt haben: ≪Seit die Mathematiker ueber die Relativitaetstheorie hergefallen sind, verstehe ich sie selbst nicht mehr!≫,' which is more accurately translated as, 'Once, in light of the mathematical finesse with which his work had been investigated and expanded, Einstein is supposed to have sighed jokingly, "Since the mathematicians have descended upon the relativity theory, even I no longer understand it."'

> of this was obscured by the absolute character of the pre-relativity time. Upon giving up the hypothesis of the absolute character of time, particularly that of simultaneity, the four-dimensionality of the time-space concept was immediately recognized. It is neither the point in space, nor the instant in time, at which something happens that has physical reality, but only the event itself. There is no absolute (independent of the space of reference) relation in space, and no absolute relation in time between two events, but there is an absolute (independent of the space of reference) relation in space and time.

Following this description, he speaks of the relativistic field as a 'four-dimensional continuum of events'; and according to this specification, everything that has or ever will happen, must exist as one absolute four-dimensional field—that which Minkowski had dubbed, in 1908, the 'absolute world'.[13]

The interpretation of reality—i.e., the ontology—Einstein thought was most objectively represented by relativity theory, is more decisively stated in *The Evolution of Physics* [193], published in 1938, which he wrote with Leopold Infeld. There, Minkowski's proposal of the absolute relativistic spacetime is clearly supported over a dynamic one. They begin their discussion by considering two different descriptions of a particle's motion in one dimension, which they claim to be the two possible objective interpretations prior to considering the results of relativity theory:

> We remember the picture of the particle changing its position with time in the one-dimensional space. We picture the motion as a sequence of events in the one-dimensional space continuum. We do not mix time and space, using a *dynamic* picture in which positions *change* with time.
>
> But we can picture the same motion in a different way. We can form a *static* picture, considering the curve in the two-dimensional time-space continuum. Now the motion is represented as something which *is*, which exists in the two-dimensional time-space contimuum, and not as something which changes in the one-dimensional space continuum.
>
> Both these pictures are exactly equivalent and preferring one to the other is merely a matter of convention and taste.
>
> Nothing that has been said here about the two pictures of the motion has anything whatever to do with the relativity theory. Both representations can be used with equal right, though classical physics favored rather the dynamical picture describing motion as happenings in space and not as existing in time-space. But the relativity theory changed this view. It was distinctly in favor of the static picture and found in this representation of motion as something existing in time-space a more convenient and more objective picture of reality. We still have to answer the question: why are these two pictures, equivalent from the point of view of classical physics, not equivalent from the point of view of the relativity theory?

[13]Minkowski even went as far as to say that the word *relativity-postulate* seemed feeble, and that, since it 'comes to mean that only the four-dimensional world in space and time is given by phenomena . . . , I prefer to call it the *postulate of the absolute world* (or briefly, the world-postulate)' [6].

> The answer will be understood if two co-ordinate systems moving uniformly, relative to each other, are again taken into account.

Then, they use the relativity of simultaneity to argue that relativity theory is only consistent with the static picture. In concluding this description of spacetime, however, they write, 'Even in the relativity theory we can still use the dynamic picture if we prefer it. But we must remember that this division into time and space has no objective meaning since time is no longer "absolute." We shall still use the "dynamic" and not the "static" language in the following pages, bearing in mind its limitations.'

In fact, we can use the dynamic description of spacetime in relativity theory because a more objective dynamic description *does* exist, with a clear representation in the theory, which could only have been considered 'less convenient' by any modern relativist because it had not been formally addressed, following Minkowski's peremptive defamation. But, in order to see that this possibility clearly must be allowed, every aspect of the problem needs to be better understood—in particular, how the relativity of simultaneity is described in that picture. But we can see already, that Einstein and Infeld's description is founded on a prior loss of the objectivity they want to claim, as they allow only the two possible descriptions, of a sequence of events in one dimension, or prior determination of a static two-dimensional spacetime continuum of events, and explicitly deny the possibility of the relative motion of two particles through a one-dimensional universe, as it moves through another dimension.

But Einstein either never did see how this could work—how a truly dynamic picture could be reconciled with the theory, as I've begun to explain—or else he, too, did not like the implications of this for the nature of reality. Apparently, he had really convinced himself of the truth of the strictly determined block universe, and held fast to that belief, to the same extent that the Eleatics had; for in Karl Popper's autobiography, he wrote that he had discussed indeterminism at length with Einstein, and 'tried to persuade him to give up his determinism, which ammounted to the view that the world was a four-dimensional block universe in which change was a human illusion, or very nearly so. (He agreed that this had been his view, and while discussing it I called him "Parmenides".)' [194]; and, in the fifth appendix to *Relativity: The Special and General Theory*, which he added in 1952, he wrote [5],

> According to the special theory of relativity . . . the sum total of events which are simultaneous with a selected event exist, it is true, in relation to a particular inertial system, but no longer independently of the choice of inertial system. The four-dimensional continuum is now no longer resolvable objectively into sections, which contain all simultaneous events; "now" loses for the spatially extended world its objective meaning. It is because of this that space and time must be regarded as a four-dimensional continuum that is objectively unresolvable, if it is desired to express the purport of objective relations without unnecessary conventional arbitrariness.
>
> . . . . The four-dimensional structure (Minkowski-space) is thought of as being the carrier of matter and of the field. Inertial spaces with their associated times are only privileged four-dimensional coordinate systems that are linked together by the linear Lorentz transformations. Since there exist in this four-dimensional structure no longer any sections which represent "now" objectively,

> the concepts of happening and becoming are indeed not completely suspended, but yet complicated. It appears therefore more natural to think of physical reality as a four-dimensional existence, instead of, as hitherto, the *evolution* of a three-dimensional existence.
>
> This rigid four-dimensional space of the special theory of relativity ...

And so finally, only weeks before Einstein's death, in a letter commemorating his friend, Michel Besso, he wrote, 'Now he has preceded me a little by parting from this strange world. This means nothing. To us believing physicists the distinction between past, present, and future has only the significance of a stubborn illusion' [16].

But while Einstein had committed the Eleatic error of not critically investigating the legitimacy of his physical principles in order to reconcile his theory with Reality, and accepted the formal implication of something which is not merely *unwelcome*, but is actually *totally preposterous*, the more common practice amongst theoretical physicists has been to take the path of disbelieving, or even entirely ignorant, sceptics, seemingly unable even to fathom the fact that the formalism we owe to Minkowski does—and was actually intended to!—describe a strictly determined block eternity, by simply ignoring that formal requirement and using the theory instead to describe paradoxical physical processes subjectively, as they occur simultaneously in any given frame.

An interesting scientific speculation which also relates to the Eleatic worldview, which was published less than a century before relativity theory, is the strict determinist theory stated in Pierre-Simon Laplace's *Essai philosophique sur les Probabilités*, which was published as an introduction to his great *Théorie analytique des Probabilités* in later editions. There, after a brief introduction, he opens the section, 'De la Probabilité', with the statement [155],

> All events, even those which on account of their insignificance do not seem to follow the great laws of nature, are a result of it just as necessarily as the revolutions of the sun. In ignorance of the ties which unite such events to the entire system of the universe, they have been made to depend upon final causes or upon hazard, according as they occur and are repeated with regularity, or appear with regard to order; but these imaginary causes have gradually receded with the widening bounds of knowledge and disappear entirely before sound philosophy, which sees in them only the expression of our ignorance of the true causes.
>
> Present events are connected with preceding ones by a tie based upon the evident principle that a thing cannot occur without a cause which produces it. This axiom, known by the name of *the principle of sufficient reason*, extends even to actions which are considered indifferent; the freest will is unable without a determinative motive to give them birth; if we assume two positions with exactly similar circumstances and find that the will is active in the one and inactive in the other, we say that its choice is an effect without a cause. It is then, says Leibniz, the blind chance of the Epicureans. The contrary opinion is an illusion of the mind, which, losing sight of the evasive reasons of the choice of the will in indifferent things, believes that choice is determined of itself without motives.
>
> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is

> animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry, added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the system of the world. Applying the same method to some other objects of its knowledge, it has succeeded in referring to general laws observed phenomena and in foreseeing those which given curcumstances ought to produce. All these efforts in the search for truth tend to lead it back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed.

The intelligence described here is now commonly called *Laplace's demon.* It is interesting to note the particular stance he has assumed regarding spacetime theory: viz., he must have thought of time as it is described in the Newtonian theory, as a disjoint, abstract dimension of physical reality; it is also clear that he has not, as Minkowski eventually did, made a final commitment to conceive of the events that would take place in time as really existing in a pre-determined state, as an eternal block spacetime; rather, he probably thought of the time-dimension, if he ever thought of it as more than a theoretical construct, as an other metaphysical dimension used to describe the events that did occur, are presently occurring, and will eventually occur, in the mode of the ancient Atomists and Aristotle, along with its formal definition in Newtonian physics.

However, given his obvious belief in the truth of an ideal God-like intelligence that could conceive of, and formulate, all of eternity, as it would necessarily be laid out when acts of free will would be relegated to the realm of infinitely complex probabilistic events, the ontological distinction between either spacetime conception becomes, in effect, what Einstein should have referred to as 'superfluous erudition'. In this way, Laplace falls into the Eleatic tradition, as a precursor of Minkowski.

But a last remark, which relates back to our discussion in § 2.4, is warranted by this over-zealous advocacy of any scientific evidence for absolute determinism: it is reasonable to assume that every physical process in Reality must obey Natural Laws which apply objectively throughout space and time, which has been done here as well; but it is nevertheless too bold an inductive leap, to suppose, from the observation of this natural order in various phenomena, that nothing should be left up to free will, which common sense tells us we must possess. Therefore, in the works of Laplace and Minkowski alike, we should recognise an excessive appeal to logical economy in discerning a mathematical description of Nature, which led each of them to propose narrow-minded theories with an unacceptable loss of objectivity, as both had, for this reason and none other, denied basic truths founded on common sense.

Chapter 3

The Dynamical Interpretation of Relativity Theory

> The point of view I am going to present does not stem from any radical view of things. I shall adopt a basically conventional attitude on most issues—or so it would have seemed, were it not for the fact that my ultimate conclusions appear to differ in their basic essentials from those most commonly expressed on this subject! My arguments do not depend on detailed calculations, but on what seem to me to be certain 'obvious' facts, whose very obviousness may contribute to their being frequently overlooked.
>
> —Sir Roger Penrose, 'Singularities and time-asymmetry'

3.1 Outline and Justification of the General Kinematical Description

According to the natural epistemology that was discussed at length in Chapter 2, our theories of nature must be considered incorrect if they cannot be reconciled with Heraclitus' principle of flux and the apparent perception of free will, or until some convincing argument, which must be consistent with common intuition, should come about that would explain how our True perceptions could be otherwise. And concerning the case of relativity theory, the goal of this chapter is to make some progress towards accomplishing such a reconciliation, by showing that it does not have to describe a block universe, as Einstein thought it did, but that it is also consistent with Heraclitus' principle.

A useful place to begin this discussion, is actually with the problem of time travel. It can be said that already by the eighteenth century A.D., the idea of a four-dimensional space-time reality, had begun to introduce itself gradually in popular literature, in the stories containing an element of time travel—although these did not usually suppose any scientific crediblilty, as most suggested a form of nonphysical time travel. However, the effect of such stories as, e.g., Charles Dickens' *A Christmas Carol* and Mark Twain's *A Connecticut Yankee in King Arthur's Court*, is no less significant for their lack of formalism, because they did serve as Sceptical illustrations of the Eleatic theory; i.e., according to which the past and future could be thought to *actually exist now*, as a real place—and, thereby being a real place, be *dynamically travelled to*, and even potentially disrupted.

In fact, it was not only the people who actually read these stories who were influenced by the ideas of time travel and four-dimensionality, any more than the theory's advance can be independently attributed to any of the authors; rather, the emergence of the theory, and its dissemination into our culture through works of science and science fiction alike, must be primarily attributed to the common growth of an idea, in the manner expressed by Arthur Koestler, when he said that [174]

> . . . the teachings of Marx and Darwin, the discoveries of Einstein and Freud, did not reach the vast majority of people in their original, printed text, but through second- and third-hand sources, through hearsay and echo. The revolutions of thought which shape the basic outlook of an age are not disseminated through text-books – they spread like epidemics, through contamination by invisible agents and innocent germ-carriers, by the most varied form of contact, or simply by breathing the common air.

But this irrational process does not limit itself to the fabrication of correct ideas; and so, even to the extent that some contribution to the theory of time travel in the twentieth century can be attributed to relativity theory, as the accepted formal theory of four-dimensional spacetime, that is not due to a correct understanding of the relativistic formalism described by Einstein, but through an irrational dissemination of the spacetime theory into popular culture, which had actually begun long before; so, too, it was not just the formal theory that eventually led some very reputable relativists to seriously investigate the time travel problem, but also their own irrational understanding of the theory's meaning, guided by an adherence to the common sense notion of Real dynamics, which the theory must strictly deny, according to its common interpretation. For it is true, as Koestler also pointed out, that 'from physics to physiology, no branch of Science, ancient or modern, can boast freedom from metaphysical bias of one kind or another' [174]—even though the centuries-long ambition of Scientists has been to achieve such a divorce. Instead, as argued in the previous chapter, the ultimate impossibility of this divorce must be accepted outright, and critical reason must be given its due place, so that we may have some control over unconscious tendencies.

So, by actually explaining, in the present chapter, how the physical time of a dynamic reality *must* be described in keeping with the relativistic formalism—i.e., how a truly dynamical metaphysics can be explained without altering the mathematical formalism of the physical theory, but only part of the commonly understood meaning,—such clear resolutions to the paradoxes of time travel and total gravitational collapse will emerge, that these will henceforth be known as characteristic examples of the 'psychological process which blinds [man] towards truths which, once perceived by a seer, become so heartbreakingly obvious'; which 'operates not only in the minds of the "ignorant and superstitious masses" as Galileo called them, but is even more strikingly evident in Galileo's own, and in other geniuses like Aristotle, Ptolemy, or Kepler', to the effect that it 'looks as if, while part of their spirit was asking for more light, another part had been crying out for more darkness' [174].

For this is in fact completely obvious in one of the more significant of the time travel stories, which did attempt a credible argument for real spacetime, in conjunction with (independently) inventing a *time machine* to provide the protagonist a means of carrying out purposeful and selective physical time travel; viz., H. G. Wells' *Time Machine*, which was published thirteen years prior to Minkowski's proposal. The novel [195] opens with its 'Time Traveller' describing his spacetime theory to a group of guests: he begins by saying that he will convince them, inasmuch as he needs them to be convinced, that the geometry they had been taught in school is incorrect; that just as a one-dimensional line, or two-dimensional plane, are not aspects of the real world, a three-dimensional, '*instantaneous* cube', that 'does not last for any time at all', is also not real. Thus (he explains to his guests), every real object must be four-dimensional, with a three-dimensional surface analogous to the well-known idea that three-dimensional objects should have two-dimensional surfaces. He argues: 'through a natural infirmity of the flesh . . . , we incline to overlook this fact,' but that 'there are really four dimensions, three which we call the three planes of Space, and a fourth, Time. There is, however, a tendency to draw an unreal distinction between the former three dimensions and the latter, because it happens that our consciousness moves intermittently in one direction along the latter from the beginning to the end of our lives;' that '*There is no difference between Time and any of the three dimensions of Space except that our consciousness moves along it.*'

He then proceeds to illustrate the four-dimensional theory: arguing that portraits are three-dimensional representations, at various points along the existences of four-dimensional beings, which are fixed and unalterable things. He introduces concepts which are reminiscent of Augustine's theory:—that scientists do make measurements of time, as in a graph of the change in barometric pressure; and that the time-dimension can be linked with our mental existences—to which he adds, by way of illustration, that we do, in our everyday lives, experience a form of time travel, when we become absent-minded, and vividly recall an incident, going back to the instant of its occurrence. 'Of course we have no means of staying back for any length of Time,' he adds, 'any more than a savage or an animal has of staying six feet above the ground. But a civilised man is better off than the savage in this respect. He can go up against gravitation in a balloon, and why should he not hope that ultimately he may be able to stop or accelerate his drift along the Time-Dimension, or even turn about and travel the other way?'

He has proceeded in graduating steps to form this idea, eventually making his argument: that, just as we can move along the two dimensions of the Earth's surface with relative ease, but have far more trouble moving up against the pull of the Earth's gravity, though we can build a balloon to take us arbitrarily in that direction, we might think also to build a vehicle capable of travelling through time. He has thus linked the flow of cosmic time with gravity, as something like the gradient of a four-dimensional gravitational field—e.g., as in the interval $r < 2m$ beyond the horizon in the Schwarzschild solution.

Despite the more inspired notion in Wells' argument, according to which a natural cause of Time is thought to be identifiable with a gravitational field, the original concept—that time is a mere illusion, and we ourselves are 'fixed and unalterable' things—is maintained; but only to the extent that those 'fixed and unalterable' things can be changed by travelling through the 'fixed and unalterable' reality in a time machine. However, the far more natural conclusion of the better part of the argument, and of general relativity theory, is the concept that in a real four-dimensional manifold, the Universe could exist as only a dynamic subspace—a dynamic hypersurface of a dynamic manifold—on which events, involving dynamic, three-dimensional bodies, take place, and are observed to do so, with different accounts of the simultaneity of past events given by relatively moving bodies within that one slice, as they also move uniformly *by necessity* through a gravitational field. Then, even if a time machine could be constructed, with the capability of going against that universal necessity, it could not take one back to a time that already happened, because 'now' and 'then' would not coexist.

Eventually, we will see more explicitly how this can naturally be described in a particular scenario; however, it is first necessary to describe how the mathematical theory of relativity—and especially the principle of relativity (i.e., moreso than the relativity of simultaneity, which we'll also get to)—is consistent with a dynamical metaphysical theory of *reality* in which every Lorentzian manifold is understood to be an *ideality*—the physical spacetime description of events—that is generally not a true block reality, but which comes to be, in the most common description, only according to a prior assumption, that in reality there is a pre-defined differential motion, or common *cosmic time*, possessed by a set of three-dimensional bodies, so that they always travel uniformly, as a particular dynamic slice of a four-dimensional gravitational field, regardless of possible relative motion *within* the universe so defined, as it evolves; i.e., that a metaphysical interpretation of relativity theory is possible, according to which the Lorentzian manifold is understood to provide the kinematical description of events that occur on a specific, well-defined dynamic hypersurface, as it moves through an equally well-defined real four-dimensional (pseudo-)Riemannian manifold, and that all bodies must remain within that evolving hypersurface *by a prior definition*, regardless of how they should rearrange themselves within that surface, relative to each other, as this well-defined, absolute cosmic time progresses.

In showing how the pre-existing relativistic formalism fully agrees with this dynamical picture, it shall become clear, once and for all, that the Universe should not really be a static, four-dimensional object, physically extended throughout spacetime, as a strictly determined block, as proponents of the Eleatic tradition continue to claim, but that it should actually be such a hypersurface evolving through a four-dimensional gravitational field. Thus, Minkowski's 'absolute world' is to be thought of, instead, as a map of all events that happen dynamically, rather than as a coagulate of deluded minds.

According to the idea of reality that is described in this interpretation of the theory, it is clear to see that the concepts of past and future should have to be re-defined, so as to formally separate the two very different senses of each—which are, e.g., confused throughout Augustine's inquiry, in the typical habit of a split-minded sleepwalker.[1] The real sense, which is being newly introduced into relativity theory, corresponds to the present universe and external void that is used to describe dynamics; whereas the sense corresponding to past and future events in the universe, which was originally promoted in status by interpreting spacetime as a real block, actually does not correspond to physical reality anywhere but at the real present, where it is conceived in the mind, or its impression is carried through reality by objects that might be perceived, although its definition comes prior to any impression that may exist at present—e.g., as described by Kant.

After defining the *present* and the *universe* as the continuously changing real coincidence

[1]From the preface to Koestler's *The Sleepwalkers* [174]: 'The history of cosmic theories, in particular, may without exaggeration be called a history of collective obsessions and controlled schizophrenias; and the manner in which some of the most important individual discoveries were arrived at reminds one more of a sleepwalker's performance than an electronic brain's.'

of matter and void, we therefore come to the required definitions: viz., *past* must refer to both:—the *ideal past*, which is the set of events in the universe that have happened already (in the dynamical sense of already); and the *real past*, which is the empty region of the four-dimensional void, through which the matter of the universe has already passed. The two senses of *future* are direct complements of those of *past*: the events in the universe that are still to come (in the dynamical sense of still), or its *ideal future*, can have had no measurement anywhere in the universe, by any relatively moving observer, as they can only be anticipated due to information that exists in the universe that occurred in its ideal past; and the *real future* of the dynamic universe is the region of the void it is always moving towards.

Note, that *reality*, as it is defined here, is technically also ideal, in the literal sense that we can form an idea, or mental image of it, according to its appropriate mathematical description; conversely, the kinematical *ideality* that is defined here, is, according to our theory, that part of the physical description of reality that is actually *un*real, but of which a clear idea nevertheless exists, as a consequence of memory, anticipation, and the finite speed of light; i.e., it is that which is in fact *purely* ideal. The language chosen here is supposed to stress the importance of this distinction: viz., that reality is *principally* real (and only *accidentally* ideal), and that ideality is *not* real.

To these definitions, it is useful to add the concept that we shall sometimes find occasion to call *now*: the dynamically evolving four-dimensional reality composed of the real past, present, and future. For example, according to the dynamical interpretation, we say that the void exists now, and that our three-dimensional Sun does not now exist in the real past, nor in the real future, but only at the present. Therefore, whereas in its proper frame in ideal spacetime it is given roughly as a four-dimensional cylinder (throughout its main-sequence lifetime), in the dynamical interpretation the Sun can be approximated, now, as a three-dimensional disc, moving through the four-dimensional void along the cosmic time dimension, in which it exists only singularly, as it also travels through the Universe.[2]

Even through this example, it should now be clear that, although the real and ideal fields must, by definition, always be equal at the present, the Lorentzian manifold cannot *generally* correspond to the geometry of a real four-dimensional dynamic void, though the actual void is described in certain specific cases; and we'll eventually see that in certain of these cases, where the real gravitational field does itself possess a prior Lorentzian signature, the cosmic time has a natural cause in the void. This fact should be kept in mind as we move on to see how the relativistic formalism is used to describe the kinematics of the dynamical picture, where it should become clear that the required prior definition of cosmic time is usually very unnaturally arbitrary—especially in special relativistic 'dynamics', where no gravitational sources exist anywhere in the void, so there are no benchmarks to which the cosmic time can be related, and therefore no means of objectively determining, at any instant in any such frame, the real present orientation of the universe on which all events in the ideal past occurred; cf., e.g., Gödel's discussion in [40].

For, whether the relativistic formalism is interpreted as such a description of dynamical events or as describing a block universe, there is actually no formal difference in the kinematical descriptions of events that occurred in any observer's past light cone—i.e., in observable

[2]More accurately, if we first define its 'radius,' then the Sun should be conceived as a dynamic agglomeration of particles in the four-dimensional void, all strictly existing within such a disc.

spacetime. In fact, in special relativity theory, the formal difference when a prior cosmic time is assumed, is that the dynamic universe really exists simultaneously in the cosmic rest-frame only—even though it is impossible to determine what that frame should be, through any objective analysis of the ideal description of events. Conversely, the block universe interpretation of relativity theory is the necessary result of imposing the unjustifiable ontological restriction that the universe must really exist simultaneously in every frame.

For it is in fact true, as Minkowski had surely realised when he briefly admitted the alternative Lorentzian interpretation of special relativity theory in 1908, that if we embrace the notion of a real 'yawning void' to the past and future, along with the prior definition of some cosmic time, the mathematical formalism is entirely consistent with the interpretation that the Universe is a three-dimensional conglomeration of three-dimensional bodies which move uniformly through a four-dimensional space; that the traditional spacetime diagram is merely the kinematical trace of this process, as it is represented in any frame, and can be represented, therefore, in any other, by introducing a change of coordinates on the spacetime; and that, when interpreted this way, the relativity of simultaneity merely means that the set of events comprising the entire graph, which occurs only on that dynamic hypersurface, as it evolves, as depicted in any frame, is perceived differently by relatively moving observers, whose perspectives differ as a result of their particular motion with respect to the causal dynamical processes which do take place—and their causal effects always remain—on that singular surface;—with the result that the particular three-dimensional set of *events* that occur *simultaneously* in the frame of an observer moving relative to the cosmic rest-frame, actually occurred in temporal sequence in reality; i.e., simultaneous phenomena in any given frame are not necessarily simultaneous noumena.

Now, in most cases in general relativity theory, the very particular way that the cosmic time needs to have been introduced *a priori*, according to the dynamical interpretation, seems very random and arbitrary, and therefore somewhat at odds with the common underlying thought that there should be no frame which is specially related to the real universe itself; but, although relativity theory requires that the descriptions of physical observations given in relatively moving frames must be equivalent, and that there is not an absolute simultaneity of perceived events, as simultaneity is perceived differently in relatively moving frames, there is actually no formal requirement to impose the reductive premiss—i.e., the further restriction, which does not strictly follow by logical deduction—that no frame should be specially related to the universe on which those events occur; and, it can similarly be said, that requiring this for every randomly moving observer, based on the fact that each gives a different account of the simultaneity of events that took place in the ideal past, unnecessarily imposes a certain 'specialness' to every such frame which subsequently *requires* an unnatural block, which must not only be strictly determined, but actually must physically exist in its entirety, as a four-dimensional block reality—a singular existence which cannot even be thought of but that it should somehow exist in some further abstract time frame, required to account for the mental delusion it still needs to describe; i.e., even the idea of the block universe is inconsistent with its singular formal description, which, in turn, is actually inconsistent with the reality it is anyhow supposed to describe.

This has also been noted by Čapek [182]:

> We shall deal only briefly with an extremely serious epistemological difficulty

> which arises when time is deprived of an ontological status and reduced to a mere appearance. For in relegating time into the phenomenal world an intolerable dualism is created between the realm of appearances, occurring in time, and the realm of timeless noumena. All static systems from Parmenides to Bradley and McTaggert are plagued by the same problem: If true reality is timeless, *where does the illusion of succession come from?* If time has no genuine reality, why does it appear to be real?
>
> No solution can be found which would not introduce surreptitiously the reality of time *somewhere*. If the illusory reality of time is nothing but a gradual rising of the curtain of ignorance which separates our mind from the complete and timeless insight, then at least *this process of rising is still a process which unfolds itself gradually without being given at once*; but, by conceding this, we admit the reality of time either in our mind or *between* our mind and the allegedly timeless reality.

Or, as Gödel remarked [40],

> It may be objected that this argument only shows that the lapse of time is something relative, which does not exclude that it is something objective; whereas idealists maintain that it is something merely imagined. A relative lapse of time, however, if any meaning at all can be given to this phrase, would certainly be something entirely different from the lapse of time in the ordinary sense, which means a change in the existing. The concept of existence, however, cannot be relativized without destroying its meaning completely. It may furthermore be different ways for different observers, whereas the lapse of time itself may nevertheless be an intrinsic (absolute) property of time or of reality. A lapse of time, however, which is not a lapse in some definite way seems to me as absurd as a coloured object which has no definite colours. But even if such a thing were conceivable, it would again be something totally different from the intuitive idea of the lapse of time, to which the idealistic assertion refers.

Formally, then, because it *is* true that the simultaneous set of events in ideal spacetime are at-once real in all relatively moving frames *if and only if* reality is a four-dimensional block universe, we must interpret the relativity of simultaneity as meaning that the simultaneous sets of ideal spacetime events of all frames are not objectively all real according to the formal description, in order to develop a theory that would describe real dynamics.

In essence, rather than assuming a real four-dimensional superposition of relative real simultaneities, and therefore strictly denying any possibility of real motion, as everyone who has followed Parmenides has done according to some infinite superposition of realities, we can accept the requirement of real dynamics that is afforded by relativity theory: that there really is an absolute three-dimensional association of three-dimensional bodies—a dynamic hypersurface of a four-dimensional relativistic field that is governed by Einstein's equation, which therefore has a unique kinematical representation in every possible frame. For it does follow, from the natural evidence against the block universe, that the interpretation that the present should really exist simultaneously in every proper frame cannot be true: whereas we have duly noted that the prior definition of a real cosmic time, given no theory to describe a natural cause, poses only an extremely serious physical problem—and we will find a natural resolution to this problem below, in agreement with general relativistic dynamics and Wells' conjecture.

As mentioned in § 2.6, this is how the dynamical interpretation of general relativity theory can resolve Zeno's paradox, 'that if place is something it must be in something' [4] (210b23). For, with time defined as a natural consequence of the four-dimensional gravitational field that is consistent with the dynamical description of the universe itself, as the natural evolution of the SdS universe actually is, according to the Jebsen-Birkhoff theorem, there is no formal requirement that the fifth dimension in the kinematical description of reality—i.e., the entire ideal evolution of *now*—should actually be real. Therefore, although the assumption of a real past and future of the present further requires an ideal past and an ideal future of now, there is no further requirement for a real past or future of now.

Now, the kinematical description of special relativity theory can actually be quite clearly given if we refer back to Wells' example of the graph of barometric pressure: although the change in pressure is traced out with time, the needle that traces out that line, as much as the line itself, and the roll of paper it is drawn on—the entire apparatus, in fact, only ever exists at the present; at the very least, it *is* only ever sensible in the present, which should mean as much. Though the line represents a passage of time, we should not say that it also exists in continuously altered fragmentary states, traced along a varying length of graph paper, throughout time, any more than we should say the rest of the apparatus, the weather that affects it, the Earth, in its progression around the Sun, etc., physically exists thoughout time, in a block; rather, we should say that the line itself exists, along with everything else, only in the dynamic present, though it holds a record of past phenomena. Similarly, when we become absent-minded, our bodies remain in the present, so that our absent-mindedness is more like the act of looking back someplace along the record of barometric pressure, rather than strictly at the needle which draws its present value, which all exists *in the present*. When we become absent-minded, we are reflecting, at present, on events that physically took place in the ideal past, which we cannot physically move to because they do not exist now, even in our real past.

An appropriate idea of a 'dynamic' special relativistic scenario can be formed—which accurately accounts for the relativity of simultaneity and 'resolves' (i.e., does not admit of) the Andromeda paradox, through its related kinematical description—by considering the graph paper in the barometer apparatus as two-dimensional Euclidean space, and the needle point as a test-particle which does not affect its structure. To begin, we'll consider that the graph paper scrolls to the left, at a rate faster than the needle is capable of moving vertically, and thus traces the pressure in a manner such that the slope of the tangent to the line that it traces out is always less than forty-five degrees, in absolute value. Next, we should add a couple more needles, so that all three are confined to exist in a verticle line: one, call it $\mathcal{A}$, positioned somewhere sufficiently below the barometer needle ($\equiv \mathcal{C}$) that the two will not touch, shall remain stationary; while the other, call it $\mathcal{B}$, located somewhere below that, will move downwards at a constant rate, say half the rate the paper scrolls to the left.

We'll also add an element to the picture, without bothering to consider the details of how it could be accomplished in this scenario, that each of these 'test-particles' is *somehow* capable of 'emitting' and 'observing' 'photons' along the one-dimensional vertical space they inhabit, which travel precisely up or down, at exactly the same rate that the paper scrolls to the left, tracing out lines along forty-five degree angles on the paper. Finally, we'll shift the relative perspective by considering the paper as being at rest, and the 'particles' as moving, instead, at an even rate to the right; but we'll retain the original coordinate system by holding that particle $\mathcal{A}$ is vertically at rest.

This is a trivial case of the dynamical scenario, involving the physical presence of test-particles confined to a hypersurface of two-dimensional Euclidean space which evolves along the greater manifold, according to a pre-defined uniform differential motion of all particles—i.e., an intrinsic cosmic time. The picture that is drawn is consistent with the relativity of simultaneity, according to the equivalent physical descriptions of the ideal past that would be given in the rest-frames of $\mathcal{A}$, $\mathcal{B}$, or any other particle in the universe moving relative to this frame with arbitrary constant velocity up to the limiting value we've set for photons—and therefore with special relativity theory.

For, as the events unfold on this two-dimensional space, an accurate special relativistic spacetime diagram is properly drawn, in this particular frame. But, it can easily be seen that each particle can, in the usual way, draw a proper coordinate basis on the chart, in which the speed of 'photons' would have the same constant value in either direction, as they had in the original frame; e.g., in $\mathcal{A}$'s frame, a straight line could be drawn through any set of events in the ideal past at a 22.5° angle to the left of vertical, which, together with $\mathcal{B}$'s worldline, would define the directions of the coordinate axes in $\mathcal{B}$'s proper frame, so that the speed of 'photons' would be said to be the same in either direction in that frame; and furthermore, the usual scaling transformation, defined by the invariant hyperbolae, should also need to be applied to each axis, so that the constant value of the speed of light would be the same in each frame.

In this new frame, all of the same rules apply to the physical description of events, including the relativity of simultaneity of events that have occurred in the ideal past, so that the laws by which physical processes are described can be stated equivalently in both. The unobservable ontological difference is that the only real physical existence that is depicted in $\mathcal{B}$'s spacetime diagram, is the one that occurs on the particular present hyperplane, which is vertical in the cosmic time rest-frame, where the particles actually *were*, together with $\mathcal{B}$, which is rotated 22.5° clockwise in $\mathcal{B}$'s proper frame; i.e., we possess *prior knowledge*, given by the way the situation was contrived, that the other two 'particles' *were* on the this particular line in the real Euclidean plane, and, in reality, *nothing* actually existed elsewhere then—in particular, if we drew the sloped line of $\mathcal{B}$'s simultaneous space in the real plane at any value of cosmic time, we know that nothing was really there; i.e., that the universe does not really exist simultaneously with $\mathcal{B}$.

In $\mathcal{B}$'s frame, where $\mathcal{B}$ is drawn as moving directly to the right, the present really evolves as a spacelike surface rotated 22.5° clockwise from the present axis of simultaneity, with the ideal past drawn to the left of the present on the spacetime diagram. In reality, however, all that exists in this frame is the angled universe, carrying information about the events that occurred in the ideal past; but, because of the Lorentz covariance in the coordinate transformation, every particle must still move uniformly to the right, regardless of any motion through the universe—and therefore, remaining in that universe, will trace out a timelike or null world-line with its tangent never having slope greater than 45° in absolute value—as no particle can, by definition, move through the universe faster than the universe moves through the underlying Euclidean void, which is well-defined in this, or any other, frame; and as the relativity of simultaneity goes: although all events occur in the universe, any two events that occur simultaneously from $\mathcal{B}$'s pespective, will not have occurred in the universe simultaneously.

Now, it is interesting to note that an interpretation, by $\mathcal{B}$ or any other relatively moving particle, that cosmic time is retarded, only follows when it is insisted that the universe really exists simultaneously in its proper frame. For it is likewise consistently true in the kinematical description, that between any two values of *cosmic time*, the elapsed value of time in $\mathcal{B}$'s frame will be less for $\mathcal{B}$, who moves along the time axis in that frame, than for an observer in the cosmic rest-frame, who moves with relative velocity, at such an angle in $\mathcal{B}$'s frame that it takes a longer coordinate time to get from one value of cosmic time to the next. In fact, in $\mathcal{B}$'s proper coordinate system, a photon emitted in the same direction that $\mathcal{B}$ is actually moving through the universe, will progress through any amount of cosmic time in less coordinate time than any other particle, while a photon emitted in the opposite direction will take longer than any other. In essence, as a result of requiring Lorentz invariance of the interval in spacetime kinematics, geometrical results such as time dilation and length contraction, and the description of a particle moving in one direction at half the speed of light, actually separating from photons emitted opposite to the direction of motion three times more rapidly than from photons emitted in that same direction, are to be thought of instead as providing a description, in any frame, that is entirely consistent with the one that is given in the cosmic rest-frame, rather than being results, due to the covariance of the transformation, that can be equivalently referenced back to ones determined in any particular other rest-frame, when the real present is assumed to be simultaneous in all such frames; for it is still true, that the amount of time that elapses on $\mathcal{A}$'s clock, in an interval of $\mathcal{B}$'s proper time, is less than the magnitude of that interval, but in the dynamical interpretation, where $\mathcal{A}$'s proper time is given as the cosmic time, the description of $\mathcal{A}$'s existence on that interval, and of $\mathcal{B}$'s existence on that interval, are not to be admitted as descriptions of two real concurrent existences.

Another important distinction between this interpretation and the block spacetime one, is made clear by qualitatively considering the randomly accelerated test-particle, $\mathcal{C}$, that was defined by the barometer's needle. In $\mathcal{C}$'s frame, spacetime must be general relativistic; but this does nothing to change the fact that the universe that is really inhabited by this particle, is the one of special relativity theory, which, in this frame, must be a curved, dynamically changing spacelike hypersurface. In fact, because the structure of spacetime in $\mathcal{C}$'s frame results partly from its randomly altering accleration through the universe (occurring in pre-defined cosmic time, on an underlying, purely Euclidean manifold), the curved present must teeter-totter in the continually changing coordinate system, as $\mathcal{C}$ moves in one direction, then the other, with respect to the cosmic rest-frame; so it should be plain to see, as a result of the inconsistency of its motion, that the only part of the underlying Euclidean reality that is actually represented in $\mathcal{C}$'s proper spacetime manifold, is the present universe itself; that its real past and future, at any instant in cosmic time, must be different from the ideal past and future that are drawn in the diagram. What this means in general, is that we must always be sure to accurately and consistently associate, with every event, the real hypersurface that is its corresponding present universe, which may be the only slice of spacetime that accurately represents reality at the same instant, if the real void would be dynamically warped as a result of the universe's mass and changing location within it.

It should already be clear, though, that in the dynamical relativity theory, there is no

Andromeda paradox; for, as the real present universe and its ideal past are consistently described in the theory, by abandoning the assumption of the reality of simultaneity, and replacing it with the covariant description of the *general relativity of universal orientation*, the problem of whether something has already happened, or is yet to be determined, depending on which relatively moving perspective it is to be considered in, does not occur.

And it must also be clear, that there is no place for any time travel paradox, such as the grandfather paradox, in our dynamical theory, since, even if real time travel, against the flow of cosmic time, could somehow be achieved, a person could not go back in time and kill their grandfather before he met their grandmother, because that ideal past does not exist now, and, as the Universe now exists—i.e., not at some instant, but in the dynamical sense of now that has been described here,—as a singular dynamic slice of all that really exists, such a time traveller could now only travel to the vacuous real past or future.

Therefore, the Andromeda paradox and the grandfather paradox clearly demonstrate two different instances of misunderstanding the metaphysical theory, due to a lack of complete acknowledgement that a strictly determined, four-dimensional corporeal continuum of events—a block universe—is formally required by the original interpretation of relativity theory—which we have replaced with a theory that clearly admits the notion of an evolving block spacetime that has recently been suggested by Ellis [28]; but, according to our theory, even the ideal past block is not real. In fact, what should now be clear, is that all of the consequences that have been derived subsequent to the picture Minkowski originally gave, when both real dynamics and the reality of simultaneity have been assumed together, should have to be re-assessed.

And so, in the dynamical interpretation, when we eventually discover an underlying natural cause for Cosmic Time, according to our cosmological theory, we shall truly see an answer to St. Augustine's conundrum [187]:

> For what is time? Who is able easily and briefly to explain that? Who is able so much as in thought to comprehend it, so as to express himself concerning it? And yet what in our usual discourse do we more familiarly and knowingly make mention of than time? And surely, we understand it well enough when we speak of it: we understand it also, when in speaking with another we hear it named. What is time then? If nobody asks me, I know: but if I were desirous to explain it to one that should ask me, plainly I know not.

For Augustine had a mixed understanding of past, present, and future; and from these, the ideal time that he commonly spoke of, was being confused with the real time that wanted further explanation; and it was this confusion that completely obscured his ability to understand the true physical aspect to which he nevertheless managed to make a significant sleepwalker's contribution. Thus, the problems of measurement of proper time, of the common sense of passing through time, and of past and future, which Augustine had thoroughly investigated, but remained considerable problems for him, as he still wanted them answered, can now be clearly understood through the formal distinction that has been made between the real and the ideal past and future, according to the dynamical idea of present, that is described by general relativistic field theory; so it becomes a simple matter to read through Augustine's entire discourse and recognise when he is speaking of reality or ideality, or even of both at once.

3.2 A Reassessment of the Potentialities

Our dynamical theory of reality, which admits a general relativistic description of physical events, and is therefore consistent with the (epistemological) principle of general relativity, has so far been described according to three physical principles: the subsistence of a real four-dimensional void; the existence, in that void, of a three-dimensional universe possessing a well-defined cosmic time; and the speed limit, within that dynamic universe, of massive particles, that is set by the constant ratio of an undisturbed photon's necessary motion through the universe, to the universe's progression through the void. However, we still want this theory to be consistent with the principle of natural cause: viz., that there has to be a basic reason for every thing which is ultimately consistent with everything else we can know of Nature, and is sufficient to explain why it should be so, and not otherwise. For, although this has led us to abandon the assumption of the reality of simultaneity, which is inconsistent with real dynamics, and although our theory actually is not *in*consistent with this principle, we have yet to discern any basic natural causes that would justify our *ad hoc* assumption of these physical principles.

Considered together, the meaning of these principles can be interpreted as a *kinematical* requirement that everything, from individual photons to galaxy clusters, must move uniformly through the four-dimensional void, but as this happens, no massive particle can move through the three-dimensional universe at the rate that a photon must. A defining characteristic of the physical description of motion therefore seems to be the null interval, which, for a special relativistic description, can be written into the underlying Euclidean void by an *ad hoc* mapping, $x_i \mapsto it$, in whichever Cartesian direction the cosmic time should be defined; and the mathematical formalism that is therefore assumed—viz., by setting $ds \equiv 0$—obscures the description of the reality it is supposed to help represent, allowing us only to compare motion of massive particles with the *ratio* of the two principal quantities of motion: the *speed* of light. Basically, by noting the consistency of photon geodesics, we define the spacetime interval *a priori* by equating (this ratio multiplied by) the cosmic time differential on the underlying void, to the differential motion of photons through the universe.

But we do recognise now, that there are two equally relevant aspects of motion in the subsistent void. We can consider each of them separately, as prior kinematical laws, without assuming the truth of the other, so long as we understand them only superficially, and don't consider the further implications for the energies of particles moving within a given frame; i.e., if we don't consider the full dynamical picture. For instance, we can require cosmic time as an absolute property of every particle in the universe, but relax the restriction on the limiting motion through the universe, so that a spaceship like the *Enterprise* could conceivably be built, that could travel to distant galaxies and back, within short intervals of cosmic time. This would not affect the fact that the spaceship would never leave the three dimensions of the universe, as required.

Alternatively, we could restore this limit on a body's rate of motion, set by the speed of light, and consider the situation in which the universe does still exist as above, as a particular hypersurface of spacetime with a natural rate of progression defining cosmic time, but relax the strictness of this progression, so that a time machine could possibly be built, to allow one to travel more (or less) rapidly through the void, than the rest of the universe, or even be allowed to reverse, and travel in the opposite direction. This would be a lonely journey, from which it would in fact be infinitely difficult to return to the universe, which is singular, and in constant motion, in that dimension.

The only hope that a time traveller could have of surviving such a journey, would be if the dynamical theory were actually incorrect, and that the block universe really exists. However, if such a time traveller did discover the truth of this, it would be of little consolation to know that the feeling of achievement in the accomplishment of time travel, and, in fact, all the wonders of life, are nothing more than the great joke of some Hand of Infinite Precision that drew an eternity and inscribed it with the property that it would fool us all, with this 'stubborn illusion', into thinking it is something completely different than it is.

Instead, we require the physical principles stated at the beginning of this section to hold absolutely. Although this might seem to be an unnatural contrivance at this point, the assumption of these principles does in fact admit the required kinematical description, and the *ad hoc* requirement of their definition, as it is so far understood to be given, shall be dealt with eventually.

To begin this, consider that our definition of cosmic time is just a formal expression of Heraclitus' principle of uniform flow, everywhere the same for all, progressing as one for all time—a truly continuously changing aspect of our existence which comes prior to the differing physical descriptions according to perspectives of relatively moving observers—the one change which precludes any motion, whence he said, 'while changing it rests'. And Heraclitus' principle is in fact so deeply natural, that if there should in fact be a fourth dimension of reality, our definition of cosmic time to describe the Heraclitean flux of the three-dimensional Universe of our perceptions, should immediately follow.

However, this is still not a sufficient *reason* for that flux to exist; it is merely a concession that the principle of cosmic time should, in some capacity, be True, if Truth would be consistent with everything we think we know of Nature. We should therefore search for a cosmological theory to describe our Universe, in which the cosmic time would actually be naturally described in a manner that is consistent with the laws of motion we've discerned in observing our dynamic Universe. The natural suspicion is, as Wells argued, that our own cosmic time might be naturally caused by the real gravitational field in which our Universe is supposed to dynamically exist—which exists, according to the field theory, in a reciprocal relation with the massive bodies it harbours, in such a manner that all things universal are required to move as one, at a well-defined rate through one of the four dimensions of the void, regardless of their mass or relative motion through the universe; i.e., that time should be described as a universal property of the gravitational field.

Thus, the motion of a body through cosmic time is supposed to be similar, in some respect, to the more intuitive notion of the motion of a mass in the spatial part of a gravitational field. For instance, we might think to find a direct analogy between the fact that the particles of the Solar wind, and the Sun itself, must move dynamically through cosmic time at the same rate, and the situation of a bullet, fired at right angles from the vertical, from some height above the surface of the Earth, and a twelve pound bowling ball, dropped from the same height at the same instant, (thus, defining the local reference frame by the time-coordinate in which the firing of the bullet and the dropping of the ball happen simultaneously), in which both very different masses, each with different velocities in the considered frame (although with the same initial velocity in the direction that matters), would fall to the Earth together, through the same gravitational potential, at the same rate.

However, if the gun were tilted towards or away from the ground, we know, respectively, that the bullet would reach the ground after a shorter or longer duration—that the two will only reach the ground at the same time if they begin with the same speed in the direction of varying gravitational potential, from precisely the same height. But we also know that the bullet, the gun, the bowling ball, and the ground, would all travel through time together, regardless of any difference in initial vertical velocity, just as the Earth and the Sun, and every photon that reaches the Earth from the Sun, are understood to move through Cosmic Time together. And if Cosmic Time should in fact naturally result from a gravitational field gradient, causing everything to move as One, it is natural to suspect that for a reason that is related to the requirement that nothing can move through space faster than light must in a vacuum, it should be impossible for any particle to be given a boost *in time*; i.e., that there is a basic dynamical relationship, or interconnectedness, between time, mass, and acceleration, which is consistently upheld in the gravitational field, which might be thought to continually mould itself according to the dynamical reciprocity so described.

For now, we will leave aside this discussion of the type of dynamic universe that we expect to describe in cosmology, and consider, instead, what right we actually have, according to relativistic kinematics, to speculate whether there should be a void fourth dimension of Reality. First of all, we may say, as the ancient Atomists realised (which was the principle of their rebuttal to the Eleatic deduction), if change is real, there must be a vessel in which this change occurs. Therefore, as we've recognised, with Heraclitus, that there is a change which subsists in all things, we already have some reason to suspect that there is a fourth dimension of a subsisting Void, through which the particles of our Universe, which variously rearrange themselves with respect to one another, also flow as One.

It is, by the way, in this assumption of a real physical dimension greater than that of our Universe, that our theory of motion formally differs from that of the ancient Atomists; but since this fourth dimension actually has no formal representation in the relativistic description of universal kinematics—i.e., because the relativistic description of the ideal evolution of the four-dimensional void pertains only to the ideal evolution of the observable present—we must in fact admit that there is really no formal requirement, based on kinematics alone, that this dimension should have the same metaphysical existence as the three-dimensions of the void inhabited by the Universe. With a pre-defined cosmic time—as our barometer illustration of the special relativistic universe signifies, and perhaps better when we think of the apparatus as being at-rest while the paper scrolls along underneath—it is possible that the physical dimension used to describe change in that cosmic time, could be a more abstract aspect of the metaphysics, with the cosmic time defined instead as an intrinsic aspect of the universe—as with Aristotle's prime mover, Augustines idea of time as a mental reach, or, in our example, the motor that turns the roll of paper along which the 'particle' world-lines are recorded—with the causal rearrangement of particles thereby occurring in an autonomous three-dimensional universe.

The difference between such an interpretation and the one that has been described above, is that in any arbitrary frame, where the present universe can be recognised (at least, once an allowable orientation in some coordinate system has been decided upon, as coinciding with the the universe and its prior definition of cosmic time), all that is said to really exist is the universe itself, possessing its intrinsic cosmic time. The past set of events will have occurred, and the future ones will eventually occur, in some yet undetermined fashion, within the universe; but there is no real past, nor real future, of the present. Even so, an arbitrary observer does not need to actually exist together with the rest of the universe *synchronously*, depending on its proper motion through the pre-defined universe; for, the universe *defined by this intrinsic abstract metaphysical change* may still be oriented, in the proper spacetime coordinates of this observer—which are still naturally assumed to describe observations, due to the finite speed of light through the universe!—exactly the same way it would be oriented if the void really extended to the past and future, and the present uniformly advanced throughout that reach, with cosmic time defined instead by a *real continuous displacement*—i.e., in the manner described in § 3.1, which leads to a physical description of events that occur in the universe that is equivalent to the one that must be given also by this continual process of rearrangement in intrinsic cosmic time.

In essence, since the fourth dimension of the void is not actually described in the physical spacetime formalism, the kinematical description that is admitted by our theory, based on the principles stated above, is exactly the same as it would be if our principles would be, instead: the subsistence of a three-dimensional void containing all matter; an abstract cosmic time, in which a particular change continually occurs that is described by a fourth physical dimension, with orthogonal orientation to a particular set of coordinates used to describe the three dimensions of the universe, which is not actually real, but 'exists' in the void only as a consequence of that prior change; and a speed limit defined by the motion of photons therein. If such were actually the case, the real fourth dimension of our theory would still be a very useful conceptual tool; but the uncomfortable fact would always remain, that within such a theory, the cosmic time can only be given as a prior *ad hoc* assumption.

It seems to me, that for the very reasons just discussed, relativity theory has always harboured Scepticism: for this is the idea of dynamics upon which relativity theory was initially based (i.e., by inherently assuming such a cosmic time, without formally defining it as such), and it is anyway entirely undesirable to have such an arbitrary definition in our theory, which cannot therefore be described by any natural cause; and even if this would be acceptable (which it is decidedly not!), it is conceptually far more difficult, when the real dynamical theory is described according to the principle in which cosmic time occurs only as an abstract change, to narrow down its place in the resulting kinematical description. And it is even more difficult to conceive of the motion of particles in a frame that is moving with respect to the cosmic rest-frame, when one can conceive only of spacetime, or else of the three-dimensional universe at an instant, and not of the universe *actually* moving uniformly through a four-dimensional void; therefore, it is very difficult to see, in such a case of real dynamism, how the relativity of simultaneity should *not* be the result of a true reality of simultaneity in every frame.[3]

So, because it would be unacceptable to have to arbitrarily assume the principle of cosmic time with no possibility of discovering a natural cause, when the reality of simultaneity really does *seem* natural enough, according to the relativistic spacetime formalism, the original interpretation of the theory was made. The caveat, however, which was probably understood

[3] As Minkowski noted, '...we are compelled to admit that it is only in four dimensions that the relations here taken under consideration reveal their inner being in full simplicity, and that on a three-dimensional space forced upon us *a priori* they cast only a very complicated projection' [6].

better by Einstein than by anyone else,[4] is that one must then strictly accept the *un*reality of dynamics, and therefore the most basic motivation for physical investigation. Relativity theory therefore met a metaphysical impasse, in that it could only correctly be thought to provide a description of either one of two naturally abhorrent theories—of which, the one, being far less comprehensible, and seeming less objective than the other,[5] was not seriously explored; and so, relativists pressed on, often describing purely dynamical processes in block universes, in the tradition of the Sceptical Empiricists who liked to assume one formalism and use it to describe the opposite, while a few fought to make sense of the theory as a formally dynamical one.

And for this same reason, our business here has been to try to identify, through logical inquiry, the natural causes that would allow for a dynamical existence described by relativity theory. To summarise the results of our argument so far:—it was necessary to concede the fact that it is actually unjustifiable to assume that the present reality always exists synchronously, and therefore completely abandon this assumption upon the ground that it is strictly inconsistent with real dynamics; but secondly, since we would in fact like to determine a natural cause for the cosmic time which must underscore the existence of things in the case of real dynamism, we've also recognised that if physical reality would be only a continually rearranging three-dimensional thing, as supposed in the Ancient atomic theory, or as stated more explicitly in the three other theories mentioned above, then our inquiry would need to end there, because it would then be impossible to determine an actual cause for that pre-defined abstract prime mover, through any theory of physical reality;[6] therefore, by deduction, the only remaining possibility that has potential to be ultimately consistent with our principle of natural cause, is the one that describes reality as consisting, now, of a four-dimensional void that is inhabited by a three-dimensional universe which moves steadily through the void according to its gravitational field law; for, if there is a real four-dimensional Void, then the entire thing should have to obey the same gravitational law as the Universe, since the matter in the Universe is supposed to move through it, and that law has apparently held consistently in the Universe throughout Time.

However, this still does not formally require Wells' conjecture, since, although the real gravitational field may continually alter as a consequence of the universe's continually changing location in it, it should not invariably cause the perpetual motion of cosmic time, which can only be required implicitly in many solutions. It is only through a further natural inclination, that we are led to suppose that our Cosmic Time might actually be an intrinsic

[4]In contrast, Eddington did prefer to think of special relativity theory as a dynamical theory, with reality 'Elsewhere' described as a complete uncertainty for all observers [13].

[5]Compare this to Einstein and Infeld's description, quoted in § 2.7, of the motion of a particle: they considered the motion either as taking place in one-dimensional space, or as being a worldline in a two-dimensional block spacetime, and did not even admit the possibility of motion within a one-dimensional space that is coherently moving through a two-dimensional void. And by completely neglecting this possibility, they concluded that the block reality was the more objective of the two options they did admit, because of the relativity of simultaneity—although it really does remain arguable whether the conclusion that all possible perceptions of simultaneity are real is *more* objective than the requirement of real dynamics—that there should be a real cosmic time, along with its corresponding cosmic rest-frame,—even when only these two options are being admitted.

[6]i.e., because no physical cause can be attributed to the existence or workings of such a supernatural entity, even if *it* is capable of exercising a physical influence on the Nature of things.

property of the supposed Void,—which, in its dynamical relationship with the Universe, must obey the gravitational field law.

Now, it is in fact because of the dynamical law that is assumed to hold consistently throughout the four-dimensional void, that we've required, from the start, that no gravitational source can exist anywhere in the void other than strictly in the dynamic universe. For if some mass would exist somewhere outside any such three-dimensional universe, our understanding of relativity theory is that it would also affect the curvature of the void, thereby influencing the evolution of the universe.

Indeed, if the Void would be dynamically warped in the presence of non-zero mass, then my own body would *now* be influencing its curvature in my immediate real past and real future, just as it would be in the space that is presently surrounding me. This poses a significant cosmological problem; for if a four-dimensional manifold truly must exist, as we've found good reason to suspect, it should then have to have existed before, or at least in conjunction with the Big Bang—and the entire system must therefore always be considered in our dynamical description. Furthermore, although the exterior manifold must not contain any sources that would influence the Universe's evolution, so that It would always progress according to Einstein's equation, it seems unsatisfactory to invoke and *ad hoc* Law of Nature requiring that voids must only contain matter in dynamic hypersurfaces. Thus, along with our search for a natural cause for cosmic time, we must in fact find a consistent cause for the existence of the void, prior to our description of the Universe. However, until that description can be given, we must necessarily continue to assume that the void is *now* perfectly vacuous everywhere but in the universe, and try formulate a consistent theory.

In particular, we should require that the duality of time, as real or ideal, should actually be described consistently by Einstein's equation, when it is applicable: for it can be used to require a geometrodynamical identity for material universes, with solutions describing the required evolution of a given universe through cosmic time—an entire universal ideality, describing the kinematics of spacetime that must take place under certain dynamical conditions, in a field which continually moulds itself according to those conditions, according to a dynamical law for the evolution of its stress-energy, as Laplace anticipated—by assuming a prior configuration of particles, which implicitly defines the cosmic time and, therefore, the actual orientation of the three-dimensional universe in the four-dimensional void; but which should also be able to be used to describe the geometry, anywhere, of the real void. For in certain vacuum solutions we do know of, these two forms of solution do coincide, so that the external vacuum of the present has the same shape it will have when the matter generating the spacetime structure is actually incident there; e.g., this is the case with the region 'outside'—i.e., in the real future of—an over-critical SdS 'mass', as well as everywhere throughout the Minkowski and de Sitter spacetimes, where the proper spacetime metrics of every inertial frame also accurately represent the real void. Therefore, we suspect that the surrounding void must obey the same vacuum field law—viz., $R_{\mu\nu} = \Lambda g_{\mu\nu}$.

But it should also be noted that certain solutions describe the dynamical evolution of a universe only, as with the RW geometries, which are completed graphs of the ideal timelines of isotropic, homogeneous, comoving universes—i.e., which really do exist simultaneously at any instant in the cosmic rest-frame, according to a prior trivial separation of space and cosmic time. But also, Einstein's equation is useful for describing test-particle kinematics in suitable local frames, even when it is understood that the supposed dynamic universes they describe, are not useful as cosmological models; e.g., in order to calculate the deflection of photons by the Sun, we can approximate it as an uncharged, non-rotating mass, and use Schwarzschild's solution, along with its implicit definition of the cosmic time in the vacuous surrounding universe, to describe the local spacetime geometry.

When this is understood, we can begin to work out how that cosmic time should be represented in Schwarzschild spacetime by considering an observer who remains at rest in the supposed infinite universe, located an infinite distance away from some Schwarzschild star, while travelling through the void, keeping the same pace as the star and any other test-particles that may also inhabit the universe, according to the causal coherence that comes with the definition of cosmic time. Since spacetime is actually Minkowskian at infinity, a natural guess is that this observer's proper time could be the cosmic time; however, according to relativistic kinematics, it is possible that the universe could be any Cauchy surface in this frame, describing simultaneity at a similar instant in the proper frame of another observer at infinity in uniform relative motion. The two important points that can already be noted, though, are that the proper time of *some particular* observer at infinity should actually be the cosmic time *there*, and that although the star's simultaneous presence is phenomenally described in all spacetime coordinate systems, the real dynamic universe in which all spacetime events occur must be only a particular spacelike slice in any frame, which transforms covariantly as in special relativity theory.

But this point is already important, because it means that one of the major motivations for belief in the hypothesis that singularities will have formed in our Universe through complete gravitational collapse—that there exist solutions of the Schwarzschild geometry which are regular at its horizon, on which the worldlines of infalling particles can be drawn so that they cross the horizon in finite 'time', while some faraway observer would 'remain' outside—should already be taken less seriously. Of course, I'm not proposing that a collapsing star must come to a dead halt at some finite radius, but, as will now be shown, that the radial coordinate singularity of Schwarzschild spacetime,

$$ds^2 = -\left(1 - \frac{R_{\text{Sch}}}{r_{\text{Sch}}}\right) dt_{\text{Sch}}{}^2 + \left(1 - \frac{R_{\text{Sch}}}{r_{\text{Sch}}}\right)^{-1} dr_{\text{Sch}}{}^2 + r_{\text{Sch}}{}^2 d\Omega^2, \tag{3.1}$$

where $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$, cannot be reached by any particle on a timelike worldline if it ever exists at any point, $(\tilde{r}_{\text{Sch}}, \tilde{t}_{\text{Sch}}) \in \{r_{\text{Sch}} > R_{\text{Sch}}, -\infty < t_{\text{Sch}} < +\infty\}$.

In fact, this is simple to prove as long as we realise, that in order for the dynamical interpretation of relativity theory to really work *the coordinates of any spacetime frame must have real metrical significance in relation to the causally coherent evolution that they describe*—for then it is possible to derive the line-element for Schwarzschild spacetime from Euclidean space, in a manner similar to our previous derivation of Minkowski spacetime, as follows: we begin by describing the observer who remains at rest at $r_{\text{Sch}} = \infty$ as travelling along a straight line, say $x_0 = x_1$, in the direction of increasing x_0 and x_1 in the skeleton geometry,

$$ds^2 = dx_0{}^2 + dx_1{}^2 + d\Omega^2 \tag{3.2}$$

(with $d\Omega^2$ defined so that it covers the 2-sphere of unit radius in $\mathbb{R}^3$), along with a three-dimensional special relativity-type universe defined by the bundle of worldlines with the same orientation, all keeping pace through the background space;[7] but rather than setting, e.g.,

$$it = \frac{x_0 + x_1}{\sqrt{2}}, \quad x = \frac{x_0 - x_1}{\sqrt{2}}, \tag{3.3}$$

(as a correct basis for describing cosmic time and the evolving universe, respectively, so that the resulting spacetime metric would provide a covariant description of the noumena that occur throughout the course of its evolution, in which null worldlines with $d\Omega = 0$ would propagate as lines of constant x_0 or x_1), we define $r_{\rm Sch} = R_{\rm Sch}$ as the limiting point, $(x_0, x_1) = (-\infty, +\infty)$, and the lines $x_0 = -\infty$ and $x_1 = +\infty$ as the lower- and upper-limits on $t_{\rm Sch}$, respectively (think of a conformal diagram); i.e., on the four-dimensional skeleton geometry, we define,

$$it_{\rm Sch} = \frac{x_0 + x_1}{\sqrt{2}}, \quad r^*_{\rm Sch} = \frac{x_0 - x_1}{\sqrt{2}}, \tag{3.4}$$

where $r^*_{\rm Sch}$ is the tortoise coordinate (cf. Box 31.2 in [41]),[8]

$$r^*_{\rm Sch} \equiv \int \frac{r_{\rm Sch}}{r_{\rm Sch} - R_{\rm Sch}} dr_{\rm Sch} = r_{\rm Sch} + R_{\rm Sch} \ln(r_{\rm Sch} - R_{\rm Sch}). \tag{3.5}$$

Then, from Eq. (3.4) we can rewrite the line-element given by Eq. (3.2) as

$$ds^2 = -dt^2_{\rm Sch} + {dr^*_{\rm Sch}}^2 + d\Omega^2, \tag{3.6}$$

which can be re-scaled by multiplying through with $dr_{\rm Sch}/dr^*_{\rm Sch} = (r_{\rm Sch} - R_{\rm Sch})/r_{\rm Sch}$:

$$\begin{aligned} \frac{dr_{\rm Sch}}{dr^*_{\rm Sch}} ds^2 &= -\frac{dr_{\rm Sch}}{dr^*_{\rm Sch}} dt^2_{\rm Sch} + \frac{dr_{\rm Sch}}{dr^*_{\rm Sch}} {dr^*_{\rm Sch}}^2 + \frac{dr_{\rm Sch}}{dr^*_{\rm Sch}} d\Omega^2 \\ &= -\frac{dr_{\rm Sch}}{dr^*_{\rm Sch}} dt^2_{\rm Sch} + \frac{dr^*_{\rm Sch}}{dr_{\rm Sch}} {dr_{\rm Sch}}^2 + \frac{dr_{\rm Sch}}{dr^*_{\rm Sch}} d\Omega^2, \end{aligned} \tag{3.7}$$

in which s and Ω can be re-scaled once more, as functions of $r_{\rm Sch} > R_{\rm Sch}$, in order to recover the metric in Schwarzschild coordinates, Eq. (3.1).

This calculation begins to illustrate the problem with arbitrarily slicing up spacetime in the usual manner, without taking care to note the metrical significance of the coordinates that are used, as well as the problem with the pseudo-Lorentzian signature of many relativistic spacetime metrics, which should often need to be synthetically contrived in the same manner; which problems will be resolved in the following section, where the most realistic means of resolving each of them will be carefully sorted out.

But two points should already be clear, in the case of Schwarzschild spacetime: $r_{\rm Sch} = R_{\rm Sch}$ is literally the end of the line, and no particle can get there along a timelike worldline.

[7] Therefore, the three-dimensional universe is defined at all times as an isotropic slice of $\mathbb{R}^4$, moving on a well-defined path in $\mathbb{R}^5$.

[8] Note that this definition sets $r^*_{\rm Sch} = r_{\rm Sch}$ at $r_{\rm Sch} = +\infty$, so that Schwarzschild radius and time coordinates do accurately describe the three-dimensional universe and cosmic time *there*, and that $r^*_{\rm Sch} = -\infty$ at $r_{\rm Sch} = R_{\rm Sch}$, so that this point does indeed correspond to $x_0 = -\infty$ and $x_1 = +\infty$ for all values of $t_{\rm Sch}$ (since $x_0/\sqrt{2} = it_{\rm Sch} + r^*_{\rm Sch}$ and $x_1/\sqrt{2} = it_{\rm Sch} - r^*_{\rm Sch}$). Also, it is useful to note that if we normalise $r_{\rm Sch}$ and $r^*_{\rm Sch}$ by $R_{\rm Sch}$, Eq. (3.5) becomes $r^*_{\rm Sch} = r_{\rm Sch} + \ln(r_{\rm Sch} - 1)$, which is actually roughly linear when $r_{\rm Sch} \not\approx 1$; and $r^*_{\rm Sch}$ does not even become negative until $r_{\rm Sch} \lesssim 1.28$.

The first point is true because $r_{\text{Sch}} = R_{\text{Sch}}$ has been *defined* as the lower-right 'corner' of the skeleton space, so that the facts that it is a positive finite value, and that r^*_{Sch} *could* be integrated as the logarithm of an absolute value function, so that aside from *right at* $r_{\text{Sch}} = R_{\text{Sch}}$ it would be well defined for all $r_{\text{Sch}} \in \mathbb{R}$, don't really matter: r^*_{Sch} is a well-behaved function right down to this limit, which is the limit of all space by prior definition. And if this weren't enough, consider as well, that the rescaling of s and Ω that would be used to recover Eq. (3.1) from Eq. (3.7), hides the fact that even if we would subsequently allow $r_{\text{Sch}} < R_{\text{Sch}}$, ignoring the actual metrical significance of the limit on r_{Sch}, the line-element changes sign on both sides of Eq. (3.7), as $r_{\text{Sch}} > R_{\text{Sch}}$ or $r_{\text{Sch}} < R_{\text{Sch}}$, so that it must still be a gross mis-interpretation to say that r_{Sch} becomes 'timelike' and t_{Sch} becomes 'spacelike' 'beyond' $r_{\text{Sch}} = R_{\text{Sch}}$, since the *basic* signature of the metric should remain as in Eq. (3.6), with that artificial definition of t_{Sch} as the 'timelike' coordinate.

The second point—that no particle that ever exists at $r_{\text{Sch}} > R_{\text{Sch}}$ can ever reach the horizon—can be understood for the following reason: lines of constant x_0 and x_1 are null lines, which describe the well-defined speed limit though the three-dimensional universe we have contrived, so that, no matter what other coordinate system might be used to describe the same spacetime (four-dimensional continuum of events), no particle that travels along a timelike worldline can ever reach $r_{\text{Sch}} = R_{\text{Sch}}$, which lies at the infinite future of the null worldline, $t_{\text{Sch}} = -\infty$; i.e., although new coordinates can be introduced, which describe synchronous slices in which some particles can reach $t_{\text{Sch}} = +\infty$ before others, none of those, when they do reach that point, will actually have come to the radial singularity of the Schwarzschild coordinate system, at $r_{\text{Sch}} = R_{\text{Sch}}$. This is because the 'point' $r_{\text{Sch}} = R_{\text{Sch}}$ is not identical to the null line $t_{\text{Sch}} = +\infty$, which is vertical in Eq. (3.2), nor the null line $t_{\text{Sch}} = -\infty$, which is horizontal in Eq. (3.2)—i.e., both are null in the $(t,\, x)$-frame defined through Eq. (3.3),—but actually exists as an infinite limit of 'space' for all t_{Sch}.

This will all be more clear when we analyse the analogous coordinate system in de Sitter space, in which all 'radial' values are finite at all 'times' other than $-\infty$ and $+\infty$, the 'radial' coordinate is indeed only real out to the horizon, and the path of the inertial observer whose worldline has all the same qualities as the one at $r_{\text{Sch}} = \infty$, whose proper time is cosmic time, etc., is in fact well-defined. For now, we conclude this examination of Schwarzschild spacetime by noting that the traditional picture of Schwarzschild collapse to a singularity at $r_{\text{Sch}} = 0$ in finite cosmological time is entirely impossible, and that the problems that have been realised as a result of that picture, such as the possible emergence of naked singularities and black hole radiation, are no more applicable in this theory of an evolving, causally coherent relativistic reality, than the prospect of time travel, so that we should no longer need to add censorship and protection conjectures to our theory.

3.3 A Reduction of Prior Theoretics

As a consequence of our examination of the Schwarzschild metric, we should like to know how, in general, mass can be incorporated into this theory. For, as opposed to the idea of a dynamical void that has been assumed so far, which reciprocally interacts with the massive bodies that move through it, the analysis of the previous section seems to correspond more closely to the description that Einstein suggested in his autobiography, when he wrote that

'If one had the field-equation of the total field, one would be compelled to demand that the particles themselves would *everywhere* be describable as singularity-free solutions of the completed field-equations. Only then would the general theory of relativity be a *complete* theory' [40].

In fact, this quotation does roughly correspond to the basic idea for the theory that we will eventually come to, and so it is useful to continue our investigation by noting that although general relativity theory describes *spacetime* as being dynamically warped in the presence of a massive body, it should not necessarily be reasonable to suppose that the geometry of the real four-dimensional void in our theory would also be warped in the common way that we think of, as that may only be the way things should be relativistically perceived in a frame in which a massive body exists.

For if the former would actually be true, we should then hope to describe the curvature of the surrounding void, in the presence of some body with *intrinsic* mass, at some instant in cosmic time, as a snapshot; and in the simplest case (if we neglect for the moment the results of our previous analysis), we might like to consider a point mass moving through the void with inertial cosmic time, and naïvely guess that it would warp the surrounding void equally in all four dimensions, and therefore simply extend the Schwarzschild solution isotropically into the fourth dimension, at some given value of cosmic time, according to

$$ds^2 \stackrel{?}{=} \frac{dr^2}{1 - 2M/r} + r^2 d{\Omega_3}^2; \tag{3.8}$$

however, this is obviously incorrect, as there are multiple issues with the possibility—e.g., notwithstanding the fact that this result does not actually follow from the Einstein equation, there is the problem that the metric signature changes at finite r.

In fact, in the actual solution, when the complete isotropy of an Einstein manifold is assumed about a given point, as we should indeed expect to be the case for the geometry surrounding any point particle with inertial cosmic time, at any value of cosmic time, mass does not simply drop out as it does in the Schwarzschild solution, where spatial isotropy and Lorentzian signature are both assumed instead; i.e., if we assume, without loss of generality, that a four-dimensional Einstein manifold is isotropic about some particular point, and therefore write its metric as[9]

$$ds^2 = A(r)dr^2 + r^2 d{\Omega_3}^2, \tag{3.9}$$

then Einstein's law, $R_{\mu\nu} = \Lambda g_{\mu\nu}$, admits two independent equations,

$$R_{rr} \quad \Rightarrow \quad \frac{3}{2}\frac{A'}{Ar} = \Lambda A, \tag{3.10}$$

$$R_{\theta\theta} \quad \Rightarrow \quad \frac{1}{2}\frac{A'r + 4A^2 - 4A}{A^2} = \Lambda r^2, \tag{3.11}$$

which can be solved for A without integration, to arrive at the actual solution,

$$ds^2 = \frac{dr^2}{1 - \frac{\Lambda}{3}r^2} + r^2 d{\Omega_3}^2, \tag{3.12}$$

[9]This choice of coordinates merely assumes the usual definition of the radial coordinate.

which is not merely isotropic about $r = 0$, but is actually a useful solution for describing all maximally symmetric real four-dimensional manifolds satisfying $R_{\mu\nu} = \Lambda g_{\mu\nu}$, which are isotropic about all points.

The complete justification for this last statement will be given shortly. For now, it is worth noting the physical consequences that may be inferred from it, due to our particular derivation. Namely, we should note that this well-known result actually does have significant meaning in our dynamical theory due to the reasoning through which it was obtained, because it can therefore be interpreted to mean: if the instantaneous four-dimensional void surrounding a particle of arbitrary 'intrinsic mass' should be isotropic, then at every instant the surrounding geometry must actually be isotropic about all points on the manifold, parametrised only by the uniform intrinsic curvature of the void; therefore, although an instant is supposed to be well-defined in our dynamical theory, an instantaneous point-mass, with purely independent existence, which would actually warp the void, thereby contributing to the dynamics of the field, is not.

The immediate implication of this, is that a particle's mass should somehow be a *consequence* of its dynamical existence, and so we find that our dynamical theory suggests a certain possibility: viz., that within our dynamical theory, there should actually be a fundamental relation between energy and cosmic time, as Hamiltonian dynamics suggest, and that a particle's mass should correspond to its motion through the universe with respect to the worldlines of photons, which are the standard of symmetry; i.e., that Einstein's result, that a massive particle must be supplied an infinite amount of energy in order to reach the speed of light, viz.

$$E = m/\sqrt{1 - v^2}, \tag{3.13}$$

should rather describe the fact that a particle, possessing energy as at exists in cosmic time, may be massive by virtue of its motion through the Universe with respect to a photon worldline; i.e., that Eq. (3.13) actually expresses the fact that photons are massless, energetic particles, also satisfying,

$$m = \sqrt{1 - v^2}E, \tag{3.14}$$

where v is the particle's velocity measured relative to some rest-frame, which is normalised by setting $c \equiv 1$. We will return to this line of inquiry below, where we shall see a connection with a statement made by Eddington, in [1], which will aid in leading to the formulation of our cosmological theory: for now, let us look more closely at the actual solution, Eq. (3.12), investigating its global properties.

To begin, we can show explicitly that the 'radii of curvature' of the maximally symmetric geometries are given by $\sqrt{3/|\Lambda|}$, by substituting $r \to r' = \sqrt{|\Lambda|/3}\, r$ into Eq.(3.12), which is valid as long as $\Lambda \neq 0$. The immediate result is,

$$ds^2 = \frac{3}{|\Lambda|}\left(\frac{dr^2}{1 - \mathrm{sgn}(\Lambda) r^2} + r^2 d{\Omega_3}^2\right), \tag{3.15}$$

where $\mathrm{sgn}(\Lambda) \equiv \Lambda/|\Lambda|$.

Another useful set of coordinates is obtained by writing

$$r = \sqrt{\frac{3}{\Lambda}} \sin\left(\sqrt{\frac{\Lambda}{3}} t\right) = \begin{cases} \sqrt{\frac{3}{\Lambda}} \sin\left(\sqrt{\frac{\Lambda}{3}} t\right), & \Lambda > 0 \\ t, & \Lambda = 0 \\ \sqrt{\frac{3}{|\Lambda|}} \sinh\left(\sqrt{\frac{|\Lambda|}{3}} t\right), & \Lambda < 0 \end{cases}, \tag{3.16}$$

so that Eq. (3.12) reduces to

$$ds^2 = dt^2 + \frac{3}{\Lambda} \sin^2\left(\sqrt{\frac{\Lambda}{3}} t\right) d\Omega_3{}^2. \tag{3.17}$$

It seems, when this result is compared with Eq. (3.15), that the maximally symmetric space, when $\Lambda > 0$, should be a 4-sphere with radius $\sqrt{3/\Lambda}$, which can be described by the 'round' metric,

$$ds^2 = \frac{3}{\Lambda}\left(d\theta^2 + \sin^2\theta d\Omega_3{}^2\right),\ 0 \leq \theta \leq \pi. \tag{3.18}$$

However, as shown below, this inference does not follow objectively from our solution, since the geometry described by Eq. (3.12) *is* continuous (although obviously not smooth, as the change in metric signature indicates) as $r \to \infty$, and the transformation, Eq. (3.16), is only valid for r out to the *coordinate singularity* at $\sqrt{3/\Lambda}$; therefore, the geometry that is actually described by Eq. (3.12), is only isometric to a *hemi*-sphere, on $r < \sqrt{3/\Lambda}$.

But we can obviously infer that, in order for this space to be maximally symmetric when $\Lambda > 0$, t must extend all the way to $\sqrt{3/\Lambda}\,\pi$, in Eqs. (3.16) – (3.17), because regardless of what Eq. (3.12) describes on $r > \sqrt{3/\Lambda}$, the geometry cannot be isotropic about all points if it consists only of the set, M_r, that is described by Eq. (3.12)—i.e., the manifold $(M_r, g(r))$,—which is only isotropic about $r < \sqrt{3/\Lambda}$ *at* $r = 0$. (Further discussion along these lines can be found in § 13.3 of [39]).

When we do allow for this inference to be made, though, we should also like to say that the metric space around every point on that sphere should be describable by a metric tensor that is equivalent to Eq. (3.12), and that the radial coordinate should extend to infinity from every such perspective, if it does so about $r = 0$, as will indeed be shown shortly; e.g., if we should centre our coordinate system about any point on the equator of the geometry described by Eq. (3.17), then $r = 0$ in Eq. (3.12) becomes a point on the horizon beyond which the new radial coordinate is supposed to extend to infinity; i.e., $r = 0$ would thus be a coordinate singularity of that new line-element.

Therefore, if Eq. (3.12) does provide a description of a positively curved manifold that is isotropic about every point, then that manifold must actually be five-dimensional, and Eq. (3.12) will describe only a particular slice of it, which is isotropic about $r = 0$; in fact, this manifold is five-dimensional de Sitter space (the five-dimensional hyperboloid of one sheet, embedded in six-dimensional Minkowski space), which will now be shown.

First of all, we can begin to see that $r = \sqrt{3/\Lambda}$ is only a coordinate singularity of the geometry that is described by our solution, Eq. (3.12), (which, we'll eventually find, is a non-differentiable critical point of the manifold defined by this particular slicing of de Sitter space, which is therefore not Riemannian itself), by writing,

$$r = \begin{cases} \sqrt{\frac{3}{\Lambda}} \sin\left(\sqrt{\frac{\Lambda}{3}} t\right), & 0 \leq t \leq \sqrt{\frac{3}{\Lambda}} \frac{\pi}{2} \\ \sqrt{\frac{3}{\Lambda}} \cosh\left(\sqrt{\frac{\Lambda}{3}} t - \frac{\pi}{2}\right), & \sqrt{\frac{3}{\Lambda}} \frac{\pi}{2} < t \end{cases}, \tag{3.19}$$

which is continuous (but only C^1) at $r = \sqrt{3/\Lambda}$. Then, the resulting metric tensor,

$$ds^2 = \begin{cases} dt^2 + \frac{3}{\Lambda} \sin^2\left(\sqrt{\frac{\Lambda}{3}} t\right) d\Omega_3{}^2, & 0 \leq t \leq \sqrt{\frac{3}{\Lambda}} \frac{\pi}{2} \\ -dt^2 + \frac{3}{\Lambda} \cosh^2\left(\sqrt{\frac{\Lambda}{3}} t - \frac{\pi}{2}\right) d\Omega_3{}^2, & \sqrt{\frac{3}{\Lambda}} \frac{\pi}{2} < t \end{cases}, \tag{3.20}$$

is no longer singular at the horizon, although we still must sort out the reason for its change in signature at that point.

In fact, this result is already quite useful, because it allows us to attribute clear geometrical meaning to Eq. (3.12), continuously for all $r > 0$, with a clear description of what goes on at $r = \sqrt{3/\Lambda}$, that gives rise to the change in metrical signature there. For, by Eqs. (3.19) – (3.20), it should be immediately apparent that the line-element describes not only half a closed sphere, when $r \leq \sqrt{3/\Lambda}$, but half of de Sitter space as well, when $r \geq \sqrt{3/\Lambda}$—i.e., since an equivalent form of the metric tensor, on the later interval, can be found from Weyl's submanifold slicing of five-dimensional Minkowski space [20, 21],

$$ds^2 = -dx_0{}^2 + \sum_{i=1}^{4} dx_i{}^2, \tag{3.21}$$

defined on the hyperboloid of one sheet,

$$-x_0{}^2 + \sum_{i=1}^{4} x_i{}^2 = \frac{3}{\Lambda} \tag{3.22}$$

(which is a more common way of describing de Sitter space), by setting

$$x_0 = \sqrt{\frac{3}{\Lambda}} \sinh\left(\sqrt{\frac{\Lambda}{3}} t\right), \tag{3.23}$$

$$x_i = \sqrt{\frac{3}{\Lambda}} \cosh\left(\sqrt{\frac{\Lambda}{3}} t\right) z_i, \quad 1 \leq i \leq 4 \tag{3.24}$$

where the z_is describe the three-sphere—i.e., $\sum_{i=1}^{4} z_i{}^2 = 1$ and $ds^2 = \sum_{i=1}^{4} dz_i{}^2$ [196]. It should be noted, however, that although the transformation given by Eqs. (3.23) – (3.24) covers the all of de Sitter space, for $-\infty < t < \infty$, the geometry that is described by Eq. (3.12) for all $r > 0$, when $\Lambda > 0$, can only be said to stitch the hemi-sphere to half of de Sitter space, joining that half, at $x_0 = 0$, continuously to the hemi-sphere's equator; and the non-differentiability of this continuous geometry at $r = \sqrt{3/\Lambda}$, as well as the change in signature of the metric tensor there, can be clearly understood, therefore, by interpreting

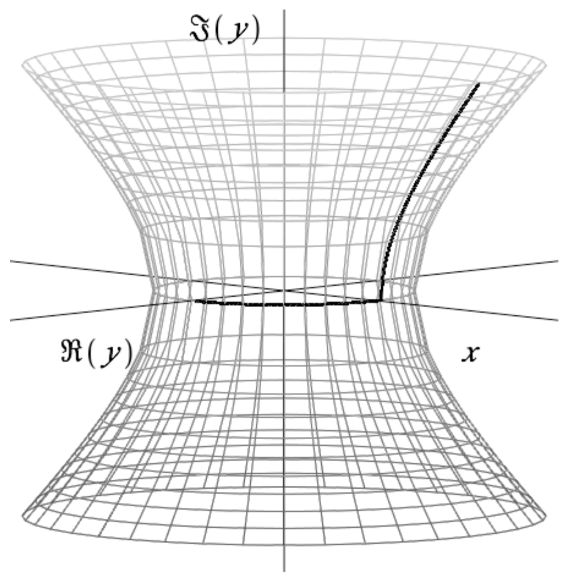

Figure 3.1: A one-dimensional analogue of the surface described by Eq. (3.12) for all $r > 0$, when $\Lambda > 0$, plotted on the two-dimensional hyperboloid of one sheet in $\mathbb{M}^3$.

the original solution, Eq. (3.12), as describing the slice of five-dimensional de Sitter space which covers half of the (positive-definite) spherical hypersurface with smallest radius (viz., half of the five-dimensional hyperboid's *equator*), along with a Lorentzian hypersurface that is bounded by that four-dimensional *hemi-sphere*'s equator; e.g., the slice of

$$ds^2 = -dt^2 + \frac{3}{\Lambda}\cosh^2\left(\sqrt{\frac{\Lambda}{3}}t\right)\left(d\theta^2 + \sin^2\theta d{\Omega_3}^2\right), \tag{3.25}$$

given by the union of the three-sphere with the set,

$$\{(t,\theta) \in \mathbb{R} :\ t = 0,\ 0 \leq \theta \leq \pi/2\ \cup\ t > 0,\ \theta = \pi/2\}. \tag{3.26}$$

We have therefore discovered, through a simple continuous rescaling of the radial coordinate, viz. Eq. (3.19), that the geometry to which Eq. (3.12) corresponds most objectively, is not '*either* the closed 4-sphere *or* four-dimensional de Sitter space, depending on the appropriate range of r', but is in fact five-dimensional de Sitter space—which may be described as a five-dimensional hyperboloid of one sheet, embedded in six-dimensional Minkowski space,—of which Eq. (3.12) gives a particular continuous (C^0) four-dimensional slicing, as follows: beginning from any point on the hyperboloid's equator, the line-element extends 90° along the equator—i.e., along the slice of five-dimensional de Sitter space in six-dimensional Minkowski space, that describes a closed 4-sphere of radius $\sqrt{3/\Lambda}$ embedded in five-dimensional Euclidean space—to reach $r = \sqrt{3/\Lambda}$, and then rotates 90° (in either direction), in order to continue along a four-dimensional slice which runs perpendicular to the plane of the equator. The one-dimensional analogue of this surface—the spherical arc, given by the function $(\mathbb{R}, \mathbb{C}, \{(x, \sqrt{\alpha^2 - x^2}) : x \in \mathbb{R} > 0\})$—is plotted in Fig. 3.1, on the two-dimensional hyperboloid in three-dimensional Minkowski space, $\mathbb{M}^3 \equiv \mathbb{R} \times \mathbb{C}$.

This is why $r = \sqrt{3/\Lambda}$ has to be called a coordinate singularity, because it is clearly not meaningful to say that the curvature goes to infinity near that point [197]; rather, this peculiar slicing is the result of our prior attempt to describe an open, positively curved Riemannian manifold, using spherical coordinates that would extend all the way in to the origin, when the geometry does not actually exist within a well-defined limit in the flat embedding space—so the line-element pivots ninety degrees when it reaches that point, and continues on to the completion of r. Therefore, when we consider the fact that the closed sphere can really be described as a subspace, of arbitrary constant radius, of any of the three other maximally symmetric spaces, it seems likely that it is only given explicit formulation in our solution, Eq. (3.12), when $\Lambda > 0$, because de Sitter space is a one-sheeted hyperboloid which has a minimum radius; so it seems far more consistent to describe de Sitter space, rather than the closed sphere, as the positive curvature analogue of the maximally symmetric spaces with zero or negative curvature, and therefore to speculate that this minimum r, as well as the Lorentzian signature, should really be intrinsic properties resulting specifically from that positive curvature. It is therefore important to examine in greater detail, the global description of the spaces with complete radial symmetry, in order to find a description that offers an objective explanation of these interesting possibilities.

In fact, if we refer back to Eqs. (3.21) – (3.22), we should note that this particular slicing of Minkowski space actually does objectively describe all three cases; e.g., when $\Lambda < 0$, Eq. (3.22) describes a hyperboloid of two sheets, which, when embedded—i.e., as a *spacelike* hypersurface—in Minkowski space, describes the negative curvature analogue of de Sitter space, which has four positive-definite eigenvalues. This can be shown explicitly, by replacing Eqs. (3.23) – (3.24) with

$$x_0 = \sqrt{\frac{3}{|\Lambda|}} \cosh\left(\sqrt{\frac{|\Lambda|}{3}} t\right), \tag{3.27}$$

$$x_i = \sqrt{\frac{3}{|\Lambda|}} \sinh\left(\sqrt{\frac{|\Lambda|}{3}} t\right) z_i, \quad 1 \le i \le 4, \tag{3.28}$$

so that Eq. (3.21) becomes instead,

$$ds^2 = dt^2 + \frac{3}{|\Lambda|} \sinh^2\left(\sqrt{\frac{|\Lambda|}{3}} t\right) d{\Omega_3}^2 \tag{3.29}$$

(cf. Eq. (3.17)).

Furthermore, it is significant that if we simply replace $x_0 \mapsto ix_0$, Eqs. (3.21) – (3.22) describe an embedding of the closed sphere of radius $\sqrt{3/\Lambda}$ in $\mathbb{R}^5$. For this reason, it is appropriate to refer to the maximally symmetric hypersurfaces of $\mathbb{M}^5$, which can all be described objectively by Eqs. (3.21) – (3.22), also as four-dimensional 'spheres' with radius $\alpha \equiv \sqrt{3/\Lambda}$ (see Figure 3.2; the formal justification for referring to all of these surfaces as 'spheres' will be made yet more explicit in what follows—especially through the analysis leading to Eq. (3.35), below).

It follows at once that geometrically distinct spherical hypersurfaces of $\mathbb{M}^5$ are described when α^2 is positive, negative, and zero: sc., when α^2 is (i.) positive, the sphere is de Sitter space—the maximally symmetric timelike hypersurface of $\mathbb{M}^5$ with positive extrinsic curvature, which is bounded according to $\sum_{i=1}^{4} {x_i}^2 = \alpha^2 + {x_0}^2 > \alpha^2$,—(ii.) negative, the sphere is

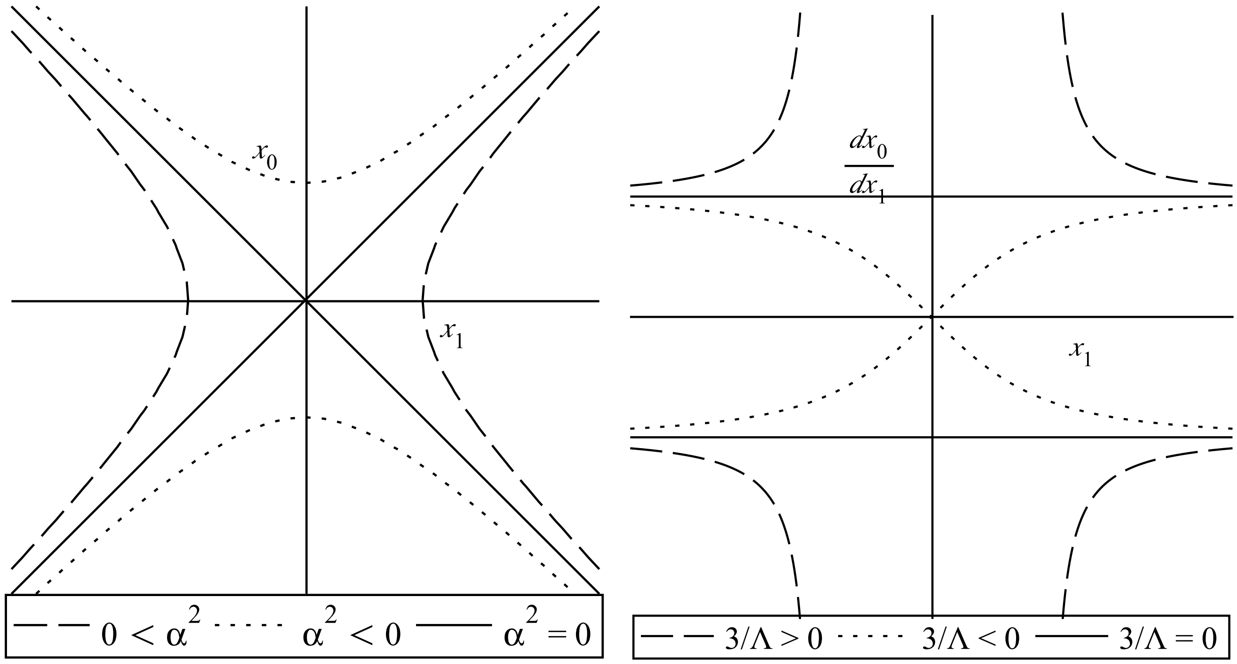

Figure 3.2: Representations, in $\mathbb{M}^2$, of the one-dimensional spheres with positive, negative, and zero radius, α (left), along with the derivative of each surface with respect to the spatial coordinate, x_1 (right). Higher-dimensional spheres are the rotations of these curves about the x_0-axis.

'Anti-de Sitter space'—the maximally symmetric spacelike hypersurface with negative extrinsic curvature, which is bounded according to $x_0^2 = -\alpha^2 + \sum_{i=1}^{4} x_i{}^2 > -\alpha^2$,—and (iii.) zero, the sphere is the null hypersurface defined by the cone, $x_0^2 = \sum_{i=1}^{4} x_i{}^2$, with zero extrinsic curvature—i.e., the zero-radius sphere, with $ds = 0$, is a (flat) light cone of the Minkowski space in which it is embedded.

Now, it must be noted that these are all (pseudo-)Riemannian manifolds, which do not need to be described as surfaces embedded in a higher-dimensional space, with extrinsic curvature—although this is useful for visually contrasting the distinct geometries,—but can in fact be described explicitly as four-dimensional Riemannian manifolds with constant *intrinsic* curvature, which all satisfy the relevant field equation,

$$R_{kilj} = \frac{1}{\alpha^2}(g_{kl}g_{ji} - g_{kj}g_{li}), \tag{3.30}$$

and therefore,

$$R_{ij} = R^k{}_{ikj} = \frac{(n-1)}{\alpha^2} g_{ij}; \tag{3.31}$$

so, although the embeddings just described have been useful for forming some intuition about each manifold, and about the differences between them, it is important to note that our analysis here has effectively amounted to an attempt at describing our assumed Void's basic state, if its curvature would be nowhere influenced by the presence of any mass, as a real four-dimensional maximally symmetric manifold.

Some recapitulation should help to clarify this significant point, to which much of our theoretical development has been converging. In § 3.2, we said that if there is actually a four-dimensional Void, then it should always have to obey the same gravitational field law as the Universe does, since the matter in the Universe is supposed to move through It and that law, which amounts to $R_{\mu\nu} \propto g_{\mu\nu}$ everywhere the stress-energy vanishes, has apparently held consistently in the Universe throughout Time. And in this section, we've been investigating the solution of this equation that describes a total vacuum which is perfectly isotropic and homogeneous. As our theoretical void goes, we've also assumed that it should subsist in Reality, regardless of any matter it might contain, and we've also assumed, in § 2.4, a principle of Natural Order—that there are basic Truths, or Natural Laws, which hold consistently throughout Physical Reality. In addition to this, we recognise the general principle of relativity theory, which requires the (empirically motivated, rationally inferred, mathematically formulated, and empirically confirmed) basic theoretical laws of physics to be a frame-independent property of Physical Reality.

But this all comes pretty well in line with an assumption that the void's basic 'ground' state should be one of the spherical solutions we've been investigating; i.e., that along with describing the vacuum spacetime geometry outside any massive source that may exist, we take Einstein's equation to be the statement of a basic physical law, that in Reality there subsists, in accordance with the Natural Order we must assume to hold throughout—so that, e.g., an Earthly electron travelling through space, say to the Coma cluster, would not become something else simply as a result of that motion, such as a stereo system pumping *Van Halen* at full blast, but would in fact remain an electron unless something occurred to alter its state of existence through some other demonstrable law of physics,—a four-dimensional Riemannian manifold with intrinsic maximal symmetry, which we call the Void.

But then the problem still remains, that aside from the description of the closed 4-sphere as a real four-dimensional hypersurface of $\mathbb{R}^5$, the rest have just been described by embeddings in an imaginary 'Minkowskian' space, $\mathbb{M}^5 \equiv \mathbb{C} \times \mathbb{R}^3$. So, let us take a step back once more, while keeping in mind the results of our discussion up to this point, and re-examine the general solution, beginning from the description of a 4-sphere with radius α, embedded in flat five-dimensional space, with line-element,

$$ds^2 = \sum_{\mu=0}^{4} dx_\mu{}^2, \tag{3.32}$$

according to the equation,

$$\sum_{\mu=0}^{4} x_\mu{}^2 = \alpha^2. \tag{3.33}$$

By solving for x_0 (without loss of generality) as a function of the other four coordinates,

$$x_0 = \pm\sqrt{\alpha^2 - \sum_{i=1}^{4} x_i{}^2}, \tag{3.34}$$

this surface can be described equivalently as a four-dimensional manifold with line-element,

$$ds^2 = d\mathbf{x}^2 + \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{\alpha^2 - \mathbf{x}^2}, \tag{3.35}$$

where $\mathbf{x} = (x_1, x_2, x_3, x_4)$ is a real four-dimensional vector in its tangent space.

In fact, Eq. (3.35), with $-\infty \leq \alpha^2 \leq \infty$, can be used to describe all maximally symmetric four-dimensional spaces, so long as appropriate ranges are assumed for the x_is, since it describes the same geometry as Eq. (3.12), in Cartesian coordinates with $\alpha^2 = 3/\Lambda$ (cf. Eq. (13.3.3) in [39], with $K = \alpha^{-2}$).

And the significance of $\mathbb{M}^5$ as an embedding space for every case other than the one for which $\alpha^2 > 0$ and $\alpha^2 > \mathbf{x}^2$, should be clear from Eq. (3.34), since x_0 must be purely imaginary otherwise, in which case Eqs. (3.32) – (3.33) can be rewritten according to $x_0 \mapsto ix_0$, so that $x_\mu \in \mathbb{R} \ \forall \ \mu \in \{0, ..., 4\}$—but according to the description given by Eq. (3.35), x_0 is not actually a coordinate of the four-dimensional manifold, which is purely real, as the x_is describing the surface are all real, regardless of the value of α^2.

Furthermore, from Eq. (3.35) it is possible to read off the corresponding elements of the metric tensor,

$$g_{ij} = \frac{1}{\alpha^2 - \mathbf{x}^2} \begin{cases} \alpha^2 - (\mathbf{x}^2 - {x_i}^2), & i = j \\ x_i x_j, & i \neq j \end{cases}, \tag{3.36}$$

which always has three positive eigenvalues, along with

$$\lambda = \frac{\alpha^2}{\alpha^2 - \mathbf{x}^2}, \tag{3.37}$$

as expected; i.e., when $\alpha^2 < 0$, $\alpha^2 = \pm\infty$ ($\Lambda = 0$), or when $\alpha^2 > 0$ and $\alpha^2 > \mathbf{x}^2$, $\lambda > 0$; when $\alpha^2 = 0$, $\lambda = 0$; and when $\mathbf{x}^2 > \alpha^2 > 0$, $\lambda < 0$.

This proves the anticipated result: that de Sitter space, with $\mathbf{x}^2 > \alpha^2 > 0$, is the only real Riemannian manifold that has maximal symmetry and intrinsic Lorentzian signature; i.e., it is a Riemannian manifold, with a real coordinate basis, which satisfies the symmetry requirement of our theory, and it is the only one of these with the requisite causal structure. In contrast, when $\lambda > 0$, the maximally symmetric spaces with positive-definite metric tensor require an *ad hoc* definition of cosmic time—e.g., which, according to our discussion in §§ 3.1 – 3.2, can be imposed on four-dimensional Euclidean space by setting $x_i = it$ for some arbitrary Cartesian coordinate, x_i, in order to impose the pseudo-Lorentzian signature of complex Minkowski space, that is required for the covariant description of spacetime events that occur in a special relativistic universe.

Now, it is an interesting consequence of the fact that de Sitter space can be described as a Lorentzian slice of Minkowski space, that although this surface has the shape of a curved hyperboloid of one sheet, the null-lines of de Sitter space must also satisfy $0 = -dx_0^2 + \sum_i dx_i^2$, and therefore must be straight lines on that curved surface in its higher-dimensional flat embedding space. This can be easily checked, e.g., by considering the equation of motion for some particle travelling along a null-line, according to Eq. (3.12):

$$\theta = \pm \int_{r > \sqrt{3/\Lambda}} \frac{dr}{r\sqrt{\frac{\Lambda}{3}r^2 - 1}} = \mp \arctan\left(\frac{1}{\sqrt{\frac{\Lambda}{3}r^2 - 1}}\right) + C, \tag{3.38}$$

which has been written in the coordinate system for which that motion occurs in only one spatial direction, θ. But note also, that in the flat Minkowski embedding space, the two-dimensional slice of de Sitter space on which this motion occurs, is the hyperboloid satisfying

$$x_0 = \pm\sqrt{{x_1}^2 + {x_2}^2 - \frac{3}{\Lambda}} = \pm\sqrt{r^2 - \frac{3}{\Lambda}}, \tag{3.39}$$

where cylindrical coordinates, for which $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$, have been introduced in the second step. Then, considering the lines for which $C = 0$ in Eq. (3.38), we find the expressions for the corresponding null-lines in Cartesian coordinates:

$$x_1 = r\cos\theta = \frac{r}{\sqrt{1 + \frac{1}{\frac{\Lambda}{3}r^2 - 1}}} = \sqrt{r^2 - \frac{3}{\Lambda}} = \pm x_0, \tag{3.40}$$

$$x_2 = r\sin\theta = \mp\frac{r}{\sqrt{\frac{\Lambda}{3}r^2 - 1}} \cdot \frac{1}{\sqrt{1 + \frac{1}{\frac{\Lambda}{3}r^2 - 1}}} = \mp\sqrt{\frac{3}{\Lambda}}, \tag{3.41}$$

as expected. So, according to Eq. (3.38), a particle travelling along a null-line will traverse an angle $\pi/2$ on the interval $\sqrt{3/\Lambda} < r < \infty$, although two lines separated by an angle π at $r = \sqrt{3/\Lambda}$ will run parallel to each other, through the flat embedding space, and will therefore never meet.

In this analysis, we have indeed found an answer to the unasked question that was at the heart of Einstein's statement, when he said, 'The non-divisibility of the four-dimensional continuum of events does not at all, however, involve the equivalence of the space co-ordinates with the time co-ordinate. On the contrary, we must remember that the time co-ordinate is defined physically wholly different from the space co-ordinates' [89]; for, rather than imposing this distinction between time and space as an abstract fact, as Einstein had to do because he believed in the relative simultaneous realities that can be described in the spacetime continuum, we now understand that this distinction is related to the requirement of a dynamic cosmic time, which can be formally described either by the *ad hoc* assumption of an absolute motion, which may be explicitly formulated through an imaginary coordinate transformation in one arbitrary direction of four-dimensional space, or it may actually be an intrinsic property of the Void, if its ground state would be four-dimensional de Sitter space.

And the fact that de Sitter space is the only maximally symmetric Riemannian manifold that possesses an intrinsic Lorentzian signature, suggests that Eddington was probably correct when he claimed that, 'to drop the cosmical constant would knock the bottom out of space' [2], which Chandrasekhar saw fit to denigrate in a centenary lecture held in his memory, when he said, 'it is clear that no serious student of relativity is likely to subscribe to [this] view' [97]; for in fact, setting $\Lambda \leq 0$ really does knock the relativistic metrical structure out of an underlying maximally symmetric void.

Actually, Eddington had stated his argument for this conclusion in more detail in [1], and it is very useful to refer back to his original reasoning, because, along with being relevant to our current analysis, it also relates back to the problem of the existence of real particles in spacetime, although some of the inferences he makes are obviously different from the ones made here; therefore, we shall consider the argument in his section dealing with the interpretation of Einstein's law of gravitation, in which a brief analysis leads him to a conclusion quite similar to ours—viz., that

…the radius of curvature in every direction and at every point in empty space has the constant length $\sqrt{3/\Lambda}$.

Conversely if the directed radius of curvature in empty space is homogeneous and isotropic Einstein's law will hold.

The statement that the radius of curvature is a constant length requires more consideration before its full significance is appreciated. Length is not absolute, and the result can only mean *constant relative to the material standards of length* used in all our measurements and in partiular in those measurements which verify $R_{\mu\nu} = \Lambda g_{\mu\nu}$. In order to make a direct comparison the material unit must be conveyed to the place and pointed in the direction of the length to be measured. It is true that we often use indirect methods avoiding actual transfer or orientation; but the justification of these indirect methods is that they give the same result as a direct comparison, and their validity depends on the truth of the fundamental laws of nature. We are here discussing the most fundamental of these laws, and to admit the validity of the indirect methods of comparison at this stage would land us in a vicious circle. Accordingly the precise statement of our result is that the radius of curvature at any point and in any direction is in constant proportion to the length of a specified material unit placed at the same point and oriented in the same direction.

This becomes more illuminating if we invert the comparison—

The length of a specified material structure bears a constant ratio to the radius of curvature of the world at the place and in the direction in which it lies. (3.42)

The law no longer appears to have any reference to the constitution of an empty continuum. It is a law of material structure showing what dimensions a specified collection of molecules must take up in order to adjust itself to equilibrium with surrounding conditions of the world.

The possibility of the existence of an electron in space is a remarkable phenomenon which we do not yet understand. The details of its structure must be determined by some unknown set of equations, which apparently admit of only two discrete solutions, the one giving a negative electron and the other a positive electron or proton. If we solve these equations to find the radius of the electron in any direction, the result must necessarily take the form

> radius of electron in given direction = numerical constant $\times$ some function of the conditions in space into which the electron was inserted.

And since the quantity on the left is a directed length, the quantity on the right must be a directed length. We have just found one directed length characteristic of the empty space in which the electron was introduced, viz. the radius of the spherical curvature of a corresponding section of the world. Presumably by going to the third or fourth derivatives of the $g_{\mu\nu}$ other independent directed lengths could be constructed; but that seems to involve an unlikely complication. There is strong ground then for anticipating that the solution of the unknown equations will be

> radius of electron in any direction = numerical constant × radius of curvature of space-time in that direction.

This leads at once to the law (3.42).

As with the electron, so with the atom and aggregations of atoms forming the practical units of material structure. Thus we see that Einstein's law of gravitation is the almost inevitable outcome of the use of material measuring-appliances for surveying the world, whatever may be the actual laws under which material structures are adjusted in equilibrium with the empty space around them.

Imagine first a world in which the curvature, referred to some chosen (non-material) standard of measurement, was not isotropic. An electron inserted in this would need to have the same anisotropy in order that it might obey the same detailed conditions of equilibrium as a symmetrical electron in an isotropic world. The same anisotropy persists in any material structure formed of these electrons. Finally when we *measure* the world, i.e. make comparisons with material structures, the anisotropy occurs on both sides of the comparison and is eliminated. Einstein's law of gravitation expresses the result of this elimination. The symmetry and homogeneity expressed by Einstein's law is not a property of the external world, but a property of the operation of measurement.

From this point of view it is inevitable that the constant Λ cannot be zero; so that empty space has a finite radius of curvature relative to familiar standards.[10] An electron could never decide how large it ought to be unless there existed some length independent of itself for it to compare itself with.

It will be noticed that our rectangular coordinates (x_1, x_2, x_3, x_4) in this and the previous section approximate to Euclidean, not Galilean, coordinates. Consequently x_4 is imaginary time, and $G_{(4)}$ is not in any real direction in this world.[11] There is no radius of curvature in a real timelike direction. This does not mean that our discussion is limited to three dimensions; it includes all directions in the four-dimensional world outside the light-cone, and applies to the space-dimensions of material structures moving with any speed up to the speed of light. The real quadric of curvature terminates at the light-cone, and the mathematical continuation of it lies not inside the cone but in directions of imaginary time which do not concern us.

By consideration of extension in timelike directions we obtain a confirmation of these views, which is, I think, not entirely fantastic. We have said that an electron would not know how large it ought to be unless there existed independent lengths in space for it to measure itself against. Similarly it would not know how long it ought to exist unless there existed a length in time for it to measure itslef against. But there is no radius of curvature in a time-like direction; so the electron does

[10] He implicitly assumes the fact that no one should consider it a real physical possibility that the radius of curvature could actually be imaginary.

[11] The errors made in this paragraph and the next one have to do with some of the points that have been discussed in our analysis so far, according to which they should be mentally corrected by the reader—for it remains relevant to us to see where Eddington was going with this.

not know how long it ought to exist. Therefore it just goes on existing indefinitely.

The alternative laws of gravitation discussed in § 62 [Alternative energy-tensors] would be obtained if the radius of the unit of material structure adjusted itself as a definite fraction not of the radius of curvature, but of other directed lengths (of a more complex origin) characteristic of empty space-time.

In § 56 [Dynamics of a particle] it was necessary to postulate that the gravitational field due to an ultimate particle of matter has symmetrical properties. This has now been justified. We have introduced a new and far-reaching principle into the relativity theory, viz. that symmetry itself can only be relative; and the particle, which so far as mechanics is concerned is to be identified with its gravitational field, is the standard of symmetry. We reach the same result if we attempt to define symmetry by the propagation of light, so that the cone $ds = 0$ is taken as the standard of symmetry. It is clear that if the locus $ds = 0$ has complete symmetry about an axis (taken as the axis of t) ds^2 must be expressible by [the most general possible solution that is spherically symmetric in space, symmetric as regards past and future time, and is static,

$$ds^2 = -U(r)dr^2 - V(r)(r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2) + W(r)dt^2, \tag{3.43}$$

where U, V, W are arbitrary functions of r.]

Our analysis has led to many conclusions which are somewhat similar to those Eddington has stated in this exerpt: viz., above, we've found that in order for the void itself, prior to the existence of mass, to intrinsically possess the Lorentzian signature that is required for general covariance, and for it to intrinsically act as a consistent measuring stick throughout space *and* time, relative to which all measurements could be objectively made, the empty void should be the de Sitter sphere, with $\Lambda > 0$. Furthermore, we've seen that photon geodesics are somehow important in the dynamical theory; therefore, it could be possible to use them as the standard of symmetry, in the way Eddington suggests. However, in contrast to Eddington's argument, our results suggest that there can be no 'ultimate particle of matter', and that the time-dimension of de Sitter space is actually as real as the dimensions of space; therefore, his speculation, that 'the electron just goes on existing indefinitely', which he based on his own understanding of both of these points, cannot be true. Instead, we've said above, that the existence of physical mass might be a property of a particle's motion relative to photon geodesics; viz., that any energetic particle which moves with respect to photon geodesics might have a measureable mass when it is considered to be at rest in spacetime, as a result of that motion.

When the results of this section are considered objectively, the possibility emerges, that if there is really no such thing as a particle with intrinsic mass, but that mass should really be a property of motion through the void, as we've found reason to suspect, then the truth may in fact be that the void does not actually dynamically warp in the presence of massive particles, but that it might actually just be de Sitter space, with the universe moving uniformly through it, and with the physical properties of matter, such as the apparent warping of spacetime around massive objects, actually corresponding to the perception of things therein, i.e. according to the principle of equivalence.

Then, the most immediate points to consider, are how the universe should actually evolve through de Sitter space, and how the paths of various particles are to be described in relation to that.

According to our analysis of de Sitter space, the most obvious possibility is that the universe could be the expanding 3-sphere that moves uniformly along r, in Eq. (3.12), beginning from a finite size at $r = \sqrt{3/\Lambda}$; or, equivalently, which evolves uniformly from a finite-sized beginning, at $t = 0$ in

$$ds^2 = -dt^2 + \frac{3}{\Lambda}\cosh^2\left(\sqrt{\frac{\Lambda}{3}}t\right) d{\Omega_3}^2; \tag{3.44}$$

and although this is indeed the description that will be given eventually, it is illustrative to first consider some other well-known solutions.

First, let us consider Eddington's statical frame, Eq. (1.5),

$$ds^2 = -\left(1 - \frac{\Lambda}{3}r_{\mathrm{Edd}}^2\right) dt_{\mathrm{Edd}}^2 + \frac{dr_{\mathrm{Edd}}^2}{1 - \frac{\Lambda}{3}r_{\mathrm{Edd}}^2} + r_{\mathrm{Edd}}^2 d\Omega^2, \tag{3.45}$$

which is found from Eqs. (3.21) – (3.22), by writing

$$x_0 = \sqrt{\frac{3}{\Lambda} - r_{\mathrm{Edd}}^2}\sinh\left(\sqrt{\frac{\Lambda}{3}}t_{\mathrm{Edd}}\right), \tag{3.46}$$

$$x_1 = \sqrt{\frac{3}{\Lambda} - r_{\mathrm{Edd}}^2}\cosh\left(\sqrt{\frac{\Lambda}{3}}t_{\mathrm{Edd}}\right), \tag{3.47}$$

$$x_i = r_{\mathrm{Edd}} z_i, \quad 2 \le i \le 4, \tag{3.48}$$

where the z_is are defined as above [196]. By using Eqs. (3.46) – (3.48), one-dimensional radial slices at constant values of t_{Edd}, as well as the t_{Edd}-evolution at various constant values of r_{Edd}, can be plotted on a two-dimensional slice of the de Sitter sphere (see Fig. 3.3).

This form of the metric describes a region of de Sitter space, symmetric about $t_{\mathrm{Edd}} = 0$, that can be described as 'the set of locally flat spacelike hypersurfaces that would be perceived, in the frame of a particle at rest at $r_{\mathrm{Edd}} = 0$, and travelling from $t_{\mathrm{Edd}} = -\infty$ to $t_{\mathrm{Edd}} = +\infty$, as being globally statical, and extending out to an observational horizon at $r_{\mathrm{Edd}} = \sqrt{3/\Lambda}$.' The prior assumption that is made, is the standard requirement that the paths of photons should be null-lines of the metric space.

For the central particle, which begins its existence at $t_{\mathrm{Edd}} = -\infty$ ($x_0 = -\infty$), it then turns out that the 'universe' at that time is also such a null-line, which extends to the horizon at $r_{\mathrm{Edd}}^2 = 3/\Lambda$ ($x_0 = 0$);[12] and the event there—or, rather, an event infinitesimally closer than that point—is the one that the central particle would observe at $t_{\mathrm{Edd}} = +\infty$, if a photon were emitted there in the appropriate direction (cf. our derivation of Schwarzschild spacetime in § 3.2).

[12]Note that the 'universe' represented by slices of constant time in Eq. (3.45), is only half as large as the slices depicted on the hyperboloid in Fig. 3.3 if it is assumed that $r_{\mathrm{Edd}} \ge 0$; i.e., because Fig. 3.3 depicts the central particle's 'surroundings' on the part of the de Sitter sphere that is covered by the coordinate system of Eq. (3.45), for $-\sqrt{3/\Lambda} \le r_{\mathrm{Edd}} \le \sqrt{3/\Lambda}$.

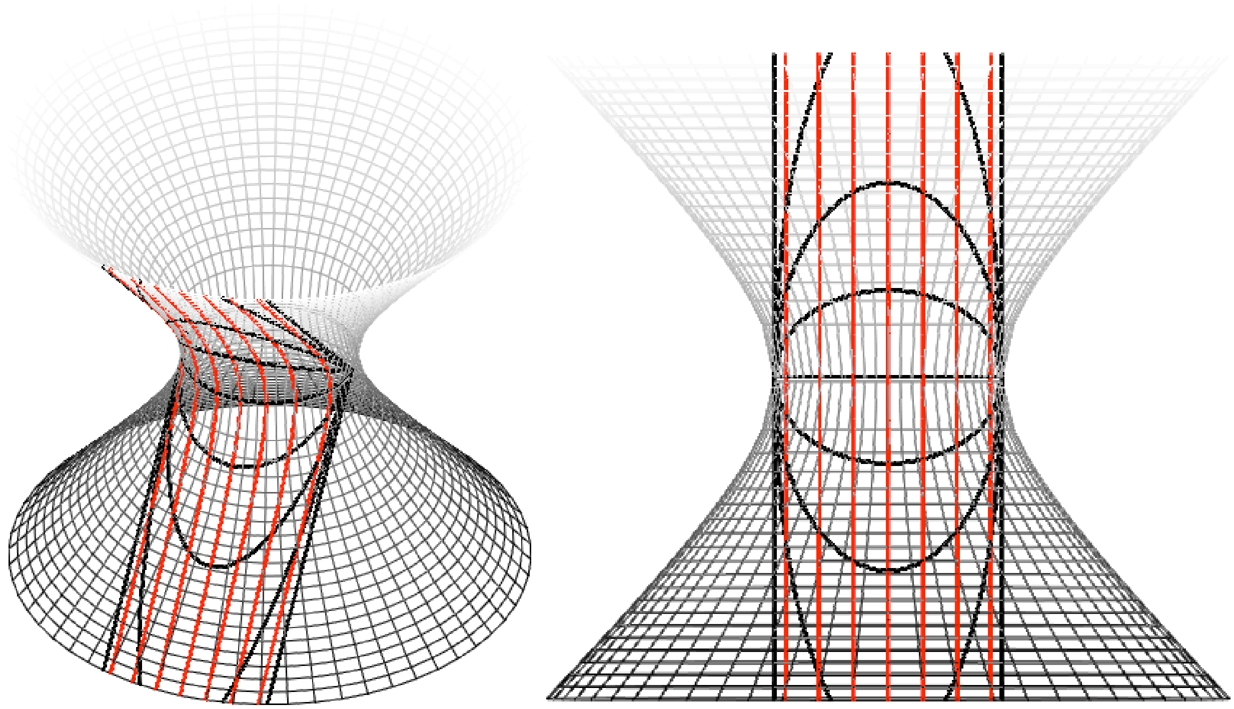

Figure 3.3: Slices of constant $t_{\rm Edd}$, for all possible real values of the radial coordinate, in Eddington's statical coordination of de Sitter space (black lines), along with (non-geodesic) world-lines of constant $r_{\rm Edd}$ (red lines), drawn on the de Sitter 2-sphere in $\mathbb{M}^3$.

Now, it should really be obvious from Fig. 3.3, that the slices of constant $t_{\rm Edd}$ in Eq. (3.45) cannot be descriptions of a dynamic universe at different values of cosmic time, because these slices are not comoving; i.e., because, according to Eqs. (3.46) – (3.48), the points in space at larger and larger values of $r_{\rm Edd}$, out to the singularity at $r_{\rm Edd} = \sqrt{3/\Lambda}$, must travel through x_0 more and more quickly until the universe reaches $x_0 = 0$, and at slower and slower rates beyond that, up to the universal limit, $r_{\rm Edd} = \sqrt{3/\Lambda}$, which reaches $x_0 = 0$ in no $t_{\rm Edd}$, and remains there forever.

In fact, it should also be clear from Fig. 3.3, or from the transformation, Eqs. (3.46) – (3.47), that if such a *dynamic* universe could possibly be, with its variable cosmic time depending on spatial location defined in order to maintain causal coherence along synchronous slices, it would *actually* be bounded at this singularity; i.e., that it is not actually correct to say that $r_{\rm Edd}$ becomes timelike and $t_{\rm Edd}$ becomes spacelike at the 'horizon' of Eq. (3.45), as we commonly used to say, based partly on the change in the metric signature, as well as the fact that the curvature does not really go to infinity there—i.e., it's really just another point of maximally symmetric de Sitter space,—but mostly because spacetime was *really* thought to be described, by relativity theory, as a four-dimensional continuum of events, something like all of de Sitter space, on which any line-element was merely thought to represent some perception of existence in which the coordinates themselves have no metrical significance.

So, we say that Eddington's coordinates really can't describe a universe in de Sitter space, as a causally coherent set of worldlines, because this slicing of maximally symmetric space is arbitrarily centred about one coordinate, with rates of timelike evolution corresponding to that arbitrary centre, and with a real spatial boundary that is required for essentially the same reasons. But the situation really only gets marginally better if we try to use the Lemaître-Robertson slicing to define the universe; for although this universe is actually comoving, and there is no longer a spatial boundary to it, it is defined by the same arbitrarily special central geodesic as Eddington's, and the causal coherence of its comoving geodesics is a gross misrepresentation of the intrinsic structure of de Sitter space.

The metric (cf. Eq. (1.8)),

$$ds^2 = -dt_{\mathrm{LR}}^2 + e^{2\sqrt{\frac{\Lambda}{3}}\,t_{\mathrm{LR}}}dy^2, \tag{3.49}$$

is found from Eqs. (3.21) – (3.22), by writing,

$$x_0 = \sqrt{\frac{3}{\Lambda}}\sinh\left(\sqrt{\frac{\Lambda}{3}}\,t_{\mathrm{LR}}\right) + \sqrt{\frac{\Lambda}{3}}\frac{r_{\mathrm{LR}}^2}{2}e^{\sqrt{\frac{\Lambda}{3}}\,t_{\mathrm{LR}}}, \tag{3.50}$$

$$x_1 = \sqrt{\frac{3}{\Lambda}}\cosh\left(\sqrt{\frac{\Lambda}{3}}\,t_{\mathrm{LR}}\right) - \sqrt{\frac{\Lambda}{3}}\frac{r_{\mathrm{LR}}^2}{2}e^{\sqrt{\frac{\Lambda}{3}}\,t_{\mathrm{LR}}}, \tag{3.51}$$

$$x_i = e^{\sqrt{\frac{\Lambda}{3}}\,t_{\mathrm{LR}}}y_i, \quad 2 \le i \le 4, \tag{3.52}$$

where $r_{\mathrm{LR}}^2 = \sum_i y_i^2$ and $dy^2 = \sum_i dy_i^2$ [196]; and, as before, Eqs. (3.50) – (3.52) have been used to plot the one-dimensional radial slices of this coordinate system, at various times, as well as the time-evolution of a few points in space, for both positive and negative values of r_{LR} (see Fig. 3.4).

The fact that the dynamic universe defined by this coordinate system is comoving, is evident from Fig. 3.4, although this is also immediately obvious from the metric, Eq. (3.49), which satisfies the requirement, that $g_{t_{\mathrm{LR}}t_{\mathrm{LR}}} = -1$, along with the requirement that constant spatial coordinates should be geodesics—i.e., that $\partial g_{it_{\mathrm{LR}}}/\partial t_{\mathrm{LR}} = 0$ [39].

This is the universe that was first described by Weyl.[13] In contrast to Eddington's coordinates, points in space represent particles in free-fall, and although different comoving geodesics move through x_0 at different rates, as in Eddington's system, the universe itself progresses uniformly in the x_0-direction, though tilted 45° from that axis, according to Eqs. (3.50) – (3.52). It is bounded by a null-line that pinches off all geodesics to a singular common origin at $x_0 = -\infty$, even though the universe, at all cosmic times thereafter, has infinite extent. In fact, it is possible to see clearly how it can be, that this universe, which is infinite in extent, also begins from a single point in de Sitter space: for although the entire bundle of geodesics emerges from a common origin at $t_{\mathrm{LR}} = -\infty = x_0$—where the spacetime metric is indeed singular,—for all $t_{\mathrm{LR}} > -\infty$ there is always a comoving geodesic connecting this point with another one at arbitrarily large x_0, or any r_{LR}, according to Eqs. (3.50) – (3.52), as illustrated in Fig. 3.4.

[13]It is fair to say that the papers by Lemaître [42] and Robertson [43] were to Weyl's analysis [85, 20, 21, 22, 23] (see also [26]; a diagram similar to Fig. 3.4 is drawn in [22]), as Martin Kruskal's [198] and George Szekeres' [199, 200] were to John Synge's [201].

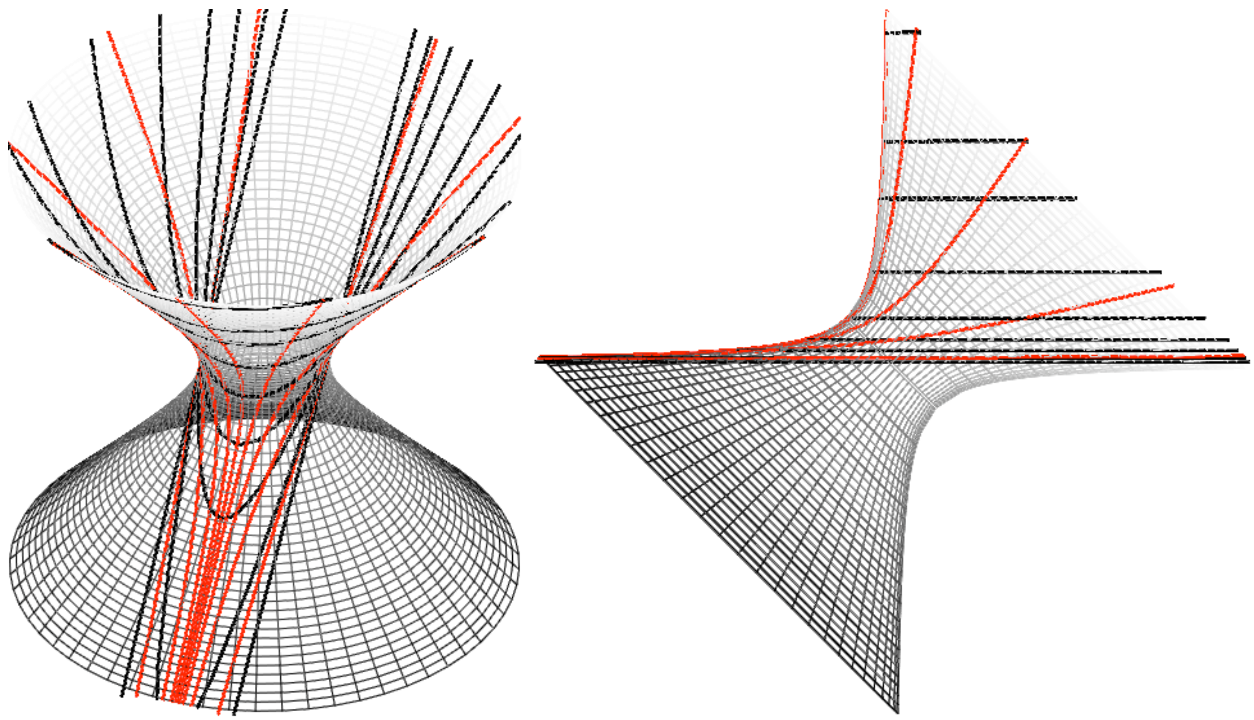

Figure 3.4: Slices of constant time in the Lemaître-Robertson coordination of de Sitter space (black lines), along with comoving geodesic world-lines (red lines), drawn on the de Sitter 2-sphere in $\mathbb{M}^3$.

However, there is still a cosmological horizon in this universe: i.e., because no information can ever come to an observer at $r_{\rm LR} = 0$ from beyond its backwards light-cone at infinite $t_{\rm LR}$, the corresponding null-line describes that observer's cosmological horizon. Now, the interesting thing about this, is that because the comoving universe is tilted 45° from the x_0-axis, particles at rest in the universe, travelling along comoving timelike geodesics, should actually be continuously streaming across that horizon for all $t_{\rm LR}$, though they can only begin to do so after having travelled an infinite distance in the x_0-direction. It is tempting to confuse this cosmological horizon—which the central particle would itself inevitably cross, if it would ever take any arbitrarily small step in the $r_{\rm LR}$-direction—with the singularity at $r_{\rm Edd} = \sqrt{3/\Lambda}$ in Eddington's coordinate system, and call that a 'removable coordinate singularity'; however, although it is true that $r_{\rm Edd} = \sqrt{3/\Lambda}$ is a singularity only of that ill-defined system, the same point of de Sitter space is not actually in the Lemaître-Robertson universe, for any $t_{\rm LR} > -\infty$.

So, the observational horizon, which is actually the same as 'space at infinite time' in Eddington's spacetime—i.e., the slice, $t_{\rm Edd} = +\infty$, of Eq. (3.45), which is always perpendicular to the comoving slices, $t_{\rm LR} = const.$, of Eq. (3.49),—has no more significance in the Lemaître-Robertson universe than that it is the horizon of the observer who remains at $r_{\rm LR} = 0$; and the fact that other particles would cross that horizon in finite cosmic time, although they would remain visible for all time to that observer, if they would have shone a light in that direction until they crossed the horizon, is ultimately due to the orientation of the evolving universe in de Sitter space. Even so, this is an interesting point, about which more will be said shortly; however, it is more prudent to first note—what is surely the most astonishing aspect of the Lemaître-Robertson universe—that although the dynamic universe, defined by Eqs. (3.50) – (3.52), is *not* actually homogeneous, it would appear so to all comoving observers.

For it is true, according to Eq. (3.49), that with photons travelling along null-geodesics, the comoving geodesics of this universe would appear to an observer at $r_{\rm LR} = 0$ as if they actually originated from flat, exponentially expanding space—and so they must also appear the same to all other comoving observers, according to the *spacetime* line-element, even though the actual dynamic universe, at any particular value of cosmic time, is an inhomogeneous slice of de Sitter space, given by Eqs. (3.50) – (3.52); i.e., although the pre-defined real present is not actually homogeneous (consider, e.g., any slice of constant $t_{\rm LR}$ in Fig. 3.4), spatial slices of the four-dimensional spacetime continuum of ideal events, at all values of comoving time in the cosmic rest-frame that is used to describe them, actually are.

But it is also true, that this particular definition of cosmic time—which requires different comoving geodesics to travel through the de Sitter sphere with varying rates depending on their distance from $r_{\rm LR} = 0$,[14] so that they would remain in causally coherent, universal connection,—makes poor use of the natural causal structure of de Sitter space that was supposed to be significant in our theory: for if we would actually consider such an arbitrarily situated universe to be plausible, it would surely not be any more reasonable a model than another universe, defined on some other Riemannian manifold by introducing its cosmic time in some other, equally arbitrary way, such as an imaginary coordinate transformation.

Furthermore, the actual inhomogeneity of the universe cannot be ignored, despite appearances; i.e., despite the fact that the spatial slices of the metric in Lemaître-Robertson coordinates *are* homogeneous for all $t_{\rm LR}$. For our stated purpose, in continuing the above investigation of maximally symmetric spaces, was that we had wanted to describe the universe as actually being homogeneous at all times, so that every place in the universe would actually be the same as every other place always. We derived our theory based on this assumption, and found that de Sitter space was the only maximally symmetric Riemannian manifold with real Lorentzian signature—i.e., the only one that intrinsically possesses those null lines which are so important in relativity theory. But this underlying homogeneity is abandoned with any slicing of the de Sitter sphere that is not actually homogeneous for all times.

With this in mind, we should consider again the fact that comoving geodesics actually do continually cross the horizon of the observer at $r_{\rm LR} = 0$—in fact, that they similarly continually cross the cosmological horizons of all comoving observers in the Lemaître-Robertson universe—by making the obvious comparison to the Schwarzschild horizon: for this is actually closely related to the theory of black holes, which was discussed at the end of § 3.2, since a similar description comes from defining a *Finkelstein universe*, as the universe with cosmic time actually given by the advanced time parameter in the Eddington-Finkelstein coordinate system, so that the Schwarzschild horizon has a similar description as the cosmological hori-

[14]As defined by Eqs. (3.50) – (3.52), comoving world-lines at larger $r_{\rm LR}$ traverse greater distances in x_0 in equal amounts of cosmic time, as graphed Fig. 3.4. In other words, the dynamic universe itself, like the Eddington universe, is not fundamentally torsionless, as it is not defined by the trivial tangent bundle of the underlying de Sitter sphere. Of course, the spacetime still has a Levi-Civita connection because it takes place on the de Sitter sphere; but the universe itself does not fundamentally coorespond to that connection.

zon in the Lemaître-Robertson universe, with the main exception being that it is the same horizon for all external observers.

But in § 3.2, it was observed that the Schwarzschild time should actually be the local measure of cosmic time at $r_{\rm Sch} = \infty$ in the pseudo-Riemannian spacetime solution that describes the evolution of a spherically symmetric star, and that the Schwarzschild metric should be completely meaningless for $r_{\rm Sch} < 2m$, where $r_{\rm Sch}$ is not simply 'timelike', but actually does not exist, just as $r_{\rm Edd} > \sqrt{3/\Lambda}$ does not exist, and the change in metric signature there is due to the fact that this spatial coordinate becomes imaginary in a metrically insignificant way. Actually, it is not that there is anything theoretically *wrong* with the Finkelstein universe, in the sense of a superficial dynamical interpretation of relativistic spacetime, aside from geodesic incompleteness at $r_{\rm Sch} = 0$: the problem is that it was not derived according to a dynamical scenario which consistently accounted for the metrical significance of the coordinates, and therefore has no real physical significance where causally coherent dynamics are concerned.

For when deriving the primary solution, we first assume, as a principle, the cosmic time, and subsequently search for a solution of Einstein's equation that would describe the isotropic motion of photons through a dynamic three-dimensional universe relative to a single point, as null lines, allowing the map of spacetime events to be coordinated differently. And even if we allow for the possibility that this velocity may depend on time, the most general possible metric is, as Eddington roughly indicated [1],

$$ds^2 = -A(r_{\rm SdS}, t_{\rm SdS})dt_{\rm SdS}^2 + B(r_{\rm SdS}, t_{\rm SdS})dr_{\rm SdS}^2 + C^2(r_{\rm SdS}, t_{\rm SdS})\left(d\theta^2 + \sin^2\theta d\phi^2\right), \qquad (3.53)$$

in which we may always write $C^2(r_{\rm SdS}, t_{\rm SdS}) \equiv r_{\rm SdS}^2$, or $C^2(r_{\rm SdS}, t_{\rm SdS}) \equiv C^2$, a constant, without loss of generality [202]. When $C^2(r_{\rm SdS}, t_{\rm SdS}) = r_{\rm SdS}^2$ and $\Lambda = 0$, one solution to the vacuum Einstein equation is the Schwarzschild solution, which time coordinate is the proper measure of cosmic time at $r_{\rm Sch} = \infty$, just as $t_{\rm Edd}$, and therefore $t_{\rm LR}$, accurately measure cosmic time at $r_{\rm Edd} = 0 = r_{\rm LR}$; cf., the derivation of Schwarzschild coordinates at the end of § 3.2. That is why the theory that gravitationally collapsing bodies have formed black holes, after having crossed their absolute event horizons in finite Cosmic Time, cannot be based on the argument that such phenomena can be described in Eddington-Finkelstein coordinates, or any other coordinate system that is principally based on Eq. (3.53), but neglects the principal dynamical scenario that was assumed.

As opposed to the paradoxical Lemaître-Robertson universe, which is somewhat similar to a 'Finkelstein universe',[15] if we now wish to describe a fundamentally evolving universe on the de Sitter sphere that really *is* homogeneous for all time, regardless of how it will be perceived according to a subsequent definition of photon worldlines, it cannot be an arbitrary slicing of de Sitter space, but must be the universe described by Eq. (3.44); i.e., the comoving sphere of radius $\sqrt{3/\Lambda}\cosh\left(\sqrt{\Lambda/3}\,t\right)$, which contracts to a finite size at $t = 0$, and exponentially

[15] Actually, a coordinate frame of de Sitter space which would really be analogous to Eddington-Finkelstein coordinates, as Eddington's statical frame is to Schwarzschild's, would retain the central geodesic, and describe the null lines at each point on that worldline as spatial slices; cf. the description in Box 31.2 of [41]. In contrast to the Eddington and Lemaître-Robertson universes, the spatial slices of such a universe would either not be smooth on $-\infty < r < \infty$, if they were chosen to be symmetric about $r = 0$ (thus resembling the absolute value function), or, if smoothness were required, they would be anti-symmetric about $r = 0$.

expands indefinitely thereafter. Recalling that the Penrose-Hawking singularity theorems require the assumption that $\Lambda \leq 0$ [203], we'll presume to eventually describe the point $t = 0$ as coinciding with the big bang in our model; therefore, when considering the cosmological theory in the next chapter, we will concentrate primarily on the half-sphere, $t \geq 0$, with the physical connection to gravitational collapse on the prior half drawn only in the final section.

Now, we are going to split from convention once more, in accordance with the theory of relativistic mass that has begun to form in this section; i.e., since we have found reason to expect that the perception of mass might be related in some way to a particle's motion. But since all particles in our dynamical theory have some motion through the void, and some particles we know of are massless, we're going to make the further guess that the motion that might relate to a particle's perceived mass is its inertial motion through the universe.

Specifically, what we're going to suppose, is that the comoving geodesics which are spatially at rest in this universe, with $d\Omega_3/dt = 0$ in Eq. (3.44), describe paths of (massless) photons—and we're going to propose to overcome the first most obvious objection to that supposition (viz., that these are comoving world-lines which isotropically and homogeneously separate from one another, and photons must be able to move in any direction through space) by saying that the cosmic rest-frame of massive particles is actually the rest-frame of a particle in the universe which moves through the de Sitter sphere along a null-line—i.e., in some direction, ϑ, through the universal 3-sphere defined by Eq. (3.44), at the rate,

$$\left(\frac{d\vartheta}{dt}\right)^2_{\text{null}} = \frac{\Lambda}{3}\cosh^{-2}\left(\sqrt{\frac{\Lambda}{3}}\,t\right), \tag{3.54}$$

—and that photons are the particles that move relative to such a particle at the same rate that *it* actually moves through the universe; therefore, we submit the requirement that photons should have *either* zero velocity through the cosmic 3-sphere *or* twice that of a null-line, and can therefore be *defined* as null-lines in the pseudo-Riemannian line-element used to describe such a general relativistic frame of rest.

In fact, we shall place no *prior* restrictions on the speed with which a particle can move through this universe.

Then, the next obvious objection to this description, is that particles should also be capable of moving in either direction through the universe, which would then imply the description of two distinct cosmic rest-frames for massive particles, corresponding to the two oppositely directed null-vectors at every point on the de Sitter sphere, and therefore three different types of photon paths, as indicated in Fig. 3.5.

A point which has to be considered before an adequate response can be made to this objection, which the line-elements we've previously considered actually attest to, is that in any 'radially' symmetric pseudo-Riemannian line-element which centres about a particle's spatial position, positive values of the 'radial' coordinate used to describe 'outer space' actually truncate at the particle, and therefore cover only half of the spacelike slice of the de Sitter sphere that the line-element describes,[16] while negative values of the 'radius' describe the particle's respective 'inner space'. Therefore, from the perspective of all 'massive particles',

[16]Namely, half of the spacelike slice through which the coordinate transformation, say from the Cartesian embedding of the real Riemannian sphere in Minkowski space, remains valid;—where it is not imaginary.

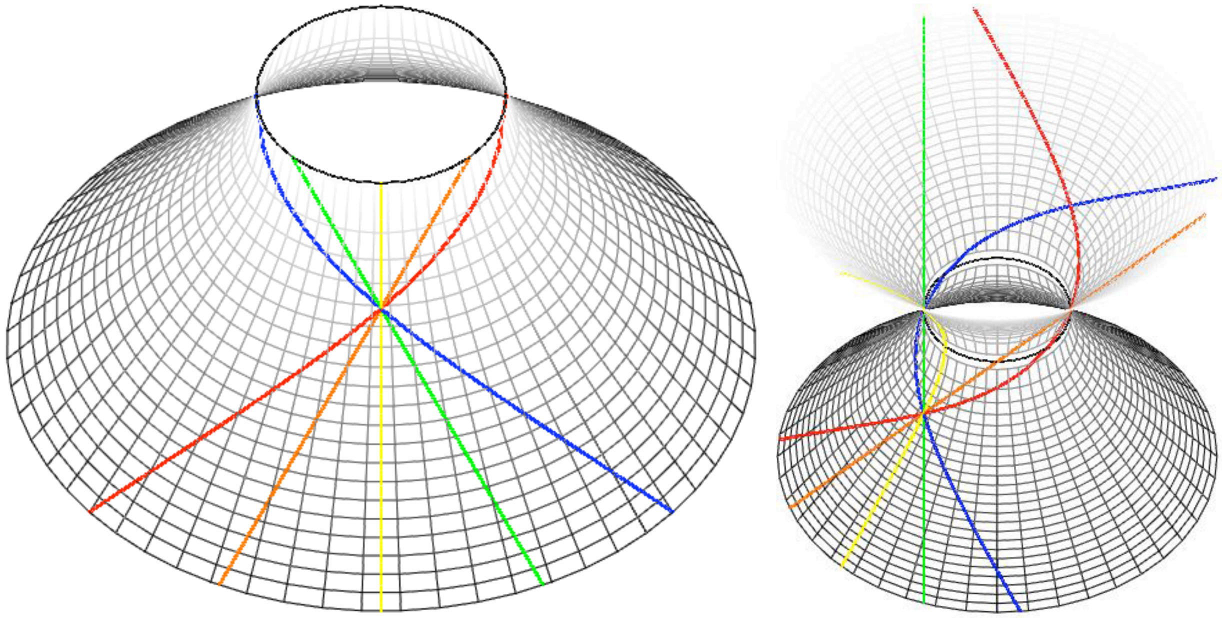

Figure 3.5: The important worldlines of our closed, comoving de Sitter cosmology, drawn on the de Sitter 2-sphere in $\mathbb{M}^3$, with t increasing downwards. The orange and green worldlines are straight null-lines which extend in either direction along the hyperboloid (doubly-ruled surface), and are therefore representative of particles that would be 'at rest' in distinct cosmic rest-frames. The other three lines, with zero velocity through the 3-sphere (yellow) or twice that of a particle moving along a null-line (red and blue), represent possible paths attributed to photons; e.g., according to a particle moving along the green worldline, the yellow and blue lines represent the paths of photons that move radially inwards and outwards, respectively.

'photons' travelling along comoving geodesics of the cosmic void, must be those which move in the negative 'radial' direction in the 'radially' symmetric spacetime defined in Eq. (3.53), while those moving at twice the null rate in the same direction as the massive particle are 'radially' outgoing; so, from the perspective of any massive particle—which we now define as one moving in any random direction, ϑ, through the universe, with any velocity other than $d\vartheta/dt = 0, \pm 2(d\vartheta/dt)_{\mathrm{null}}$ (which correspond, due to general covariance, to massless particles travelling along null-lines in *its* proper line-element, whether it is inertial ($d\vartheta/dt = const.$) or not)—another particle that moves through the universe in the opposite direction would be 'moving radially inward through space faster than a photon'. Particles on oppositely directed orbits are therefore very different species.

But what is really meant by 'oppositely directed orbits through the evolving 3-sphere'? In the two-dimensional representation given in Fig. 3.5, which graphs the cosmic evolution of a 1-sphere, this is quite easy to see, since there are two objectively well-defined directions around a circle; but it is not so clear as we come to consider the timelike evolution of a 3-sphere. For example, in the case of a 2-sphere it is possible to parallel transport a tangent vector along a path that would bring it back to the same point with the opposite (or any other) sense.[17] In fact, the 2-sphere is not parallelisable, according to the hairy ball theorem [204], so 'directions of motion' on it can't be completely well-defined. It is therefore remarkable that the 3-sphere is parallelisable, as a 2-sphere worth of 1-spheres with a twist in the manifold [205], so that two universal 'opposite directions of motion' can always be defined, as with the evloving 1-sphere depicted in Fig. 3.5.

So, further to the assumption that the cosmographical line-element should be parametrised by the evolution of massive particles travelling along null-lines on the de Sitter sphere, so that the paths of massless particles satisfying $d\vartheta/dt = 0$, which actually move relative to a rest-frame observer at the same rate that the observer moves through the universe, can therefore be *defined* as null in the corresponding spacetime geometry, we recognise that the direction of motion of these fundamental worldlines can be defined by any inertial particle according to a universal parallelisation.

And the gist of the idea, as it seems to apply according to the discussion in § 4.2, is that an inertial particle's velocity through the 3-sphere should be able to be related, in the local radially symmetric coordinate frame, to some sort of *radial* momentum, even though those coordinates describe the particle as remaining at the origin—and the two 'opposite directions of motion' that we are now considering, are therefore opposite 'radial' directions of motion (i.e., relative 'inwards' and 'outwards', corresponding to positive or negative radial velocity), which is not described as taking place *along* the radial coordinate in the proper frame of such a particle (i.e., a particle does not describe itself as 'expanding' or 'contracting'), but is still described by the metric, Eq. (3.53), in a different way, because such particles are actually moving through the real dynamic universe along inertial paths which are not geodesics.

What seems to be an interesting consequence of the fact that there are three distinct velocities for photons, is that the mass of any particle is not a single value, but a triplet—an objectively well-defined superposition of three values, which seems to indicate a basic relation between our theory of relativity and something like de Broglie-Bohm theory—and the physical difference we shall find, between particles that do move with opposite velocities through the universe, is that their masses should therefore be each other's negatives.

But because we are primarily concerned with describing the cosmic evolution from the perspective of a collection of 'positive' mass particles, we can begin by considering the motion of all particles through the universe as being all with the same sense of direction, corresponding to one set of null-lines on the de Sitter sphere. When this is done, it is actually quite simple to sort out the most general form for the line-element, even without using a coordinate transformation from any standard de Sitter line-element, since Eddington's symmetry idea applies similarly in this case; viz., we take the homogeneous three-dimensional universe to be moving 'radially' outwards in the cosmic time dimension, but re-define one spatial coordinate to meet the requirement, that a *photon* moving in either direction in that dimension, follows a null-line—i.e.,

$$ds^2 = -B(r_{\rm SdS}, t_{\rm SdS})dr_{\rm SdS}^2 + A(r_{\rm SdS}, t_{\rm SdS})dt_{\rm SdS}^2 + r_{\rm SdS}^2\left(d\theta^2 + \sin^2\theta d\phi^2\right). \tag{3.55}$$

We have thus performed a speed-time contraction on the de Sitter sphere [44] which preserves the metrical description of the flow of cosmic time that is given in the comoving coordinates

[17] For example, transport a tangent vector 180° along the great circle in the direction that it is pointing, and then 180° along a perpendicular one, so that it arrives back at its original position with opposite sense.

of Eq. (3.44); but rather than describing space according to the inertial geodesics of that universe, and preserving the basic null-structure of the void, the spatial coordinates are described according to the null worldlines that all move in the same direction with respect to a parallelisation of the universe, relative to which the (unaccelerated) inertial *geodesics* really do move at the cosmic null rate, so that they may reasonably be defined as null lines in the contracted line-element. The result of solving the vacuum Einstein equation for this general line-element has a well-known form, similar to the solution when A and B have the opposite signs, as in Eq. (3.53); in addition to those due to its 'radial' symmetry, the metric may also be given with a t_{SdS}-like Killing vector, so that a particle's velocity through space does not depend on position; and the inhomogeneity in the spatial coordinate t_{SdS} merely reflects the inversion of perspective that has been written into the line-element, according to which that spatial coordinate should now run parallel to the oppositely-directed null-lines of the real de Sitter void.

In the following chapter, we will analyse the cosmology entirely from this inverted perspective, beginning from a derivation of the SdS geometry, as the simple solution to the vacuum Einstein equation, from this perspective, that describes the cosmic evolution of this universe along r_{SdS}. And it will eventually be shown, that from this perspective the universe would be perceived as a massive one that began from a singularity at $r_{\mathrm{SdS}} = 0$, in such a manner that, from the cosmic rest-frame, space would be perceived as being flat, with the universe expanding from this finite beginning, and scaling according to $r_{\mathrm{SdS}}(\tau) \propto \sinh^{2/3}\left(\frac{3}{2}\sqrt{\frac{\Lambda}{3}}\tau\right)$—the precise scale-factor of the flat ΛCDM model.

Chapter 4

A New Cosmology

> We have introduced a new and far-reaching principle into the relativity theory, viz. that symmetry itself can only be relative; and the particle, which so far as mechanics is concerned is to be identified with its gravitational field, is the standard of symmetry.
>
> —Sir Arthur Stanley Eddington, *The Mathematical Theory of Relativity*

4.1 Global Parametrisation of Spacetime in a Massive Rest-Frame

According to the discussion in Chapter 3, we shall proceed here with the analysis of a comoving 3-sphere which expands from a finite minimum radius that is well-defined in its subsisting de Sitter void, based on the additional prior physical assumption, that the cosmic rest-frame of massive particles *is* the collection of null-lines emanating in one direction through that void, with 'massless' photons defined as the particles which move at the same rate relative to this fundamental rest-frame—i.e., either zero velocity through the actual universe, or twice that of a null line,—so that the spacetime paths of massive particles and photons may be described by defining photon geodesics as null rays in the line-element corresponding to the perception of events in that frame, given by Eq. (3.55).[1] And the analysis in this chapter will eventually justify the subsequent assertion that was made there, that such a universe would actually appear to all observers, if they would model the universe's large-scale evolution using the proper line-element in the cosmic rest-frame, precisely as a flat ΛCDM universe that began from a singularity at the beginning of cosmic time.

In order to justify that assertion, however, it is not enough to consider only the model that would be used to describe the observed cosmic expansion rate, but we must also work out what the large-scale structure of all matter in the universe should be, and how it would appear to massive observers. We will come to that description through a determination of the physical parametrisation of the spacetime geometry which applies not only to the cosmic rest-frame, but similarly to the geometry that describes any local massive frame which may

[1] In this chapter, we will drop the subscript from r_{SdS} and t_{SdS}, which are the primary such coordinates analysed here. They are not to be confused with the r and t coordinates defined in Eqs. (3.9) and (3.19).

be considered on a large enough scale that overall charge and spin should be negligible, which might instead be given by the alternate form of the metric, Eq. (3.53). After the physical parametrisation of the geometries and the cosmographical theory have been completed in this section and the next, we will proceed, in § 4.3, to determine the line-element in the proper time coordinate system of the fundamental rest-frame, and the corresponding cosmological model. Finally, in § 4.4, we shall conclude with the qualitative description of a scenario for the birth of such a universe following from gravitational collapse in a similar prior one.

Thus, we begin by determining possible restrictions on the coefficients $A(r,t)$ and $B(r,t)$ in Eq. (3.55), when the metric tensor is required to satisfy the field equation, $R_{\mu\nu} = \Lambda g_{\mu\nu}$; in particular, we immediately note that

$$0 = R_{rt} = R_{tr} = -\frac{1}{rB(r,t)}\frac{\partial B(r,t)}{\partial t}. \tag{4.1}$$

Therefore, $B(r,t)$ must be a function of r only, which we make use of in order to simplify the expressions for the remaining components of the Ricci tensor ($X' \equiv \partial X/\partial r$):

$$R_{rr} = -\frac{B}{4A'}\frac{\partial}{\partial r}\left(\frac{A'^2}{AB}\right) + \frac{B'}{Br}, \tag{4.2}$$

$$R_{tt} = \frac{A}{4A'}\frac{\partial}{\partial r}\left(\frac{A'^2}{AB}\right) + \frac{A'}{Br}, \tag{4.3}$$

$$R_{\theta\theta} = 1 + \frac{1}{B} - \frac{B'r}{2B^2} + \frac{A'r}{2AB}, \tag{4.4}$$

$$R_{\phi\phi} = R_{\theta\theta}\sin^2\theta; \tag{4.5}$$

so that, according to the vacuum Einstein equation, we may write,

$$\frac{R_{rr}}{B} + \frac{R_{tt}}{A} = \frac{(AB)'}{AB^2r} = 0. \tag{4.6}$$

Therefore, $(AB)' = 0$, and we may generally write $A(r,t) = f(t)/B(r)$; but then $f(t)$ may be gauged away by re-defining t through $dt \mapsto dt' = \sqrt{f(t)}dt$, so that

$$A(r) = \frac{1}{B(r)}, \tag{4.7}$$

which generalises the Jebsen-Birkhoff theorem [36, 37, 38] to non-zero Λ, as proven by Lemaître in 1931 [73], according to which the metric of spherically symmetric spacetime beyond a non-charged mass may always be written with an additional t-like Killing vector field. For a more detailed analysis, see [206], which generalises the statement of the theorem and its proof to include the Nariai solution [207, 208, 209, 210],

$$ds^2 = -\left(1 - \Lambda r^2\right)dt^2 + \frac{dr^2}{1 - \Lambda r^2} + \frac{1}{\Lambda}\left(d\theta^2 + \sin^2\theta d\phi^2\right). \tag{4.8}$$

Now, we also have the requirement that $R_{\theta\theta} = \Lambda r^2$; thus, insertion of Eq. (4.7) into Eq. (4.4) yields

$$1 + \frac{1}{B} - \frac{B'r}{B^2} = 1 + \frac{d}{dr}\left(\frac{r}{B}\right) = \Lambda r^2, \tag{4.9}$$

or

$$1 + A + A'r = 1 + \frac{d}{dr}(Ar) = \Lambda r^2; \tag{4.10}$$

either of which may be directly integrated, with the result,

$$A(r) = \frac{1}{B(r)} = \frac{\frac{\Lambda}{3}r^3 - r - \left(\frac{\Lambda}{3}{r_0}^3 - r_0\right) + A(r_0)r_0}{r}, \tag{4.11}$$

where r_0 is some limit of the coordinate r.

So far, no constraint has been given on r, other than the requirement of its definition, that hyper-surfaces of constant r and t will have an area $4\pi r^2$, which defines it as 'the radius', and allows us to set $g_{\theta\theta} \equiv r^2$. In fact, this does not even require r to be positive. In order to make this definition, however, $r = 0$ needs to be relinquished as a coordinate singularity of the metric. Actually, when the integration constant in Eq. (4.11),

$$2M \equiv r_0 - \frac{\Lambda}{3}{r_0}^3 + A(r_0)r_0, \tag{4.12}$$

is not zero, $r = 0$ is the well-known black hole singularity (cf. Eq. (1.6)). Integration of Eqs. (4.9) and (4.10) is therefore not valid across this value.

Furthermore, we should observe that r_0 is a horizon of the geometry if, and only if, $A(r_0) = 1/B(r_0) = 0$. Then, it is a coordinate singularity through which we also may not integrate, and, as such, an appropriate reference point from which to parametrise the solution; we therefore set $A(r_0) \equiv 0$, in order to provide physical meaning to the other integration constant, r_0.

But also, since $\Lambda > 0$ by definition on our underlying void, it is extremely useful to write the solution in a scale-invariant form, normalising all dimensional quantities with the universal 'quadric of curvature', as Eddington called it, according to $s \mapsto s' = \sqrt{\Lambda/3}s$, $t \mapsto t' = \sqrt{\Lambda/3}t$, $r \mapsto r' = \sqrt{\Lambda/3}r$, and $r_0 \mapsto {r_0}' = \sqrt{\Lambda/3}r_0$ ($\Rightarrow M \mapsto M' = \sqrt{\Lambda/3}M$), so that the metric can finally be written,

$$ds^2 = -\frac{r}{r^3 - {r_0}^3 - (r - r_0)}dr^2 + \frac{r^3 - {r_0}^3 - (r - r_0)}{r}dt^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{4.13}$$

Obviously, this solution always has the one horizon, at $r = r_0$, as our definition of r_0 requires; but when ${r_0}^2 \leq 4/3$ there are actually two more, at $r = -\frac{r_0}{2} \pm \frac{1}{2}\sqrt{-3r_0^2 + 4}$;—i.e., when this condition is met, there are three real solutions to the horizon equation,

$$r^3 - {r_0}^3 - (r - r_0) = 0, \tag{4.14}$$

and the two horizons that exist in addition to the one which is always explicitly defined at $r = r_0$, are on the ellipse,

$$r^2 + rr_0 + {r_0}^2 = 1. \tag{4.15}$$

When ${r_0}^2 < 4/3$ *and* ${r_0}^2 \neq 1/3$ (i.e., $M^2 < 1/27$), and the line-element has three distinct singularities in addition to the one at $r = 0$, positive r is spacelike on some interval, and the solution is in the form given by Eq. (3.53), with $C(r, t) \equiv r^2$, which describes spacetime surrounding an uncharged, non-spinning mass, in which radial photons move isotropically along null-lines. This shall be referred to as the *under-critical* SdS geometry.

When the condition, $r_0^2 = 4/3$ or $1/3$ ($M^2 = 1/27$), for the *critical* SdS geometry is met, there are precisely two horizons: one at $r = r_0$, and another at $r = -r_0/2$ or $-2r_0$, as the case may be—i.e., the two horizons of the two critical SdS geometries are at $r = \pm 2/\sqrt{3}$ and $\mp 1/\sqrt{3}$, with corresponding mass parameter, $M = \mp 1/\sqrt{27}$.

And when ${r_0}^2 > 4/3$ ($M^2 > 1/27$), there is only one real root of Eq. (4.14), and the sign of M differs from that of r_0. The line-element of this *over-critical* SdS geometry, which is timelike for all $r > 0$ when $M > 1/\sqrt{27}$ ($r_0 < -2/\sqrt{3}$), provides an ideal spacetime description of our universe, which begins from an initial singularity at $r = 0$, and expands as the cosmic time, r, increases. But $r = 0$ is very different from the *big bang* singularities of the dynamical FLRW universes which must emerge in some manner like the Einstein-de Sitter universe, so that, as Eddington put it, 'in the beginning all the matter created was projected with a radial motion so as to disperse even faster than the present rate of dispersal of galaxies' [2]—for this property is common to the initial states of all such universes, since the description at the beginning is not essentially different even if some 'dark energy'—even a quintessence field that could be linked to the inflation scenario—should eventually take over as the dominant cause of expansion. Instead, the SdS singularity at $r = 0$ is of the de Sitter-type, and therefore much like that of the Lemaître-Robertson universe, about which Weyl wrote, 'If I think about how, on the de Sitter hyperboloid the world lines of a star system with a common asymptote rise up from the infinite past, then I would like to say: the World is born from the eternal repose of 'Father Æther'; but once disturbed by the 'Spirit of Unrest' (*Hölderlin*), which is at home in the Agent of Matter, 'in the breast of the Earth and Man', it will never come again to rest' [22]; so the mathematical description of cosmic expansion is in fact consistent with Eddington's interpretation.

The global parametrisation of the SdS geometries is graphed in Fig. 4.1. In the shaded regions r is 'timelike', and in the unshaded regions it is 'spacelike'. The radius at which the gravitational potential vanishes identically, $r = \sqrt[3]{-2M} \equiv \sqrt[3]{{r_0}^3 - r_0}$, has been plotted along with two other radii of interest, $r = 2M$ and $r = 3M$. It is important to note, that the geometry for any particular value of r_0, when ${r_0}^2 \leq 4/3$, is precisely the same as when r_0 has instead the value of either of the other two horizon radii, as it must be, according to the explicit definition of the cubic *mass parameter*, M. In other words, the geometries in the parametric range, $-2/\sqrt{3} \leq r_0 \leq -1$, are formally equivalent to those in the intervals, $0 \leq r_0 \leq 1/\sqrt{3}$ and $1/\sqrt{3} \leq r_0 \leq 1$; as well, they are reflections about $r = 0$, of the geometries for which $-1 \leq r_0 \leq -1/\sqrt{3}$, $-1/\sqrt{3} \leq r_0 \leq 0$, and $1 \leq r_0 \leq 2/\sqrt{3}$, which are similarly equivalent. There is a degeneracy between the three values of the horizons, requiring that any given value of the mass parameter, which is used superficially to describe a given geometry, must be the same if r_0 would be any of the other two corresponding values in the under-critical parameter range. In this case, we find that the three solutions to Eq. (4.14) form one triplet—for a given value of any one unambiguously fixes those of the other two, according to the cubic equation that defines the mass parameter. But in the two over-critical ranges, at $r_0 < -2/\sqrt{3}$ and $r_0 > 2/\sqrt{3}$, which are reflections of each other just as the two formally distinct critical and under-critical geometries are, no such triplet of real values of r_0 exists.

As noted in the derivation of the metric, Eq. (4.13), there is no analytical necessity for

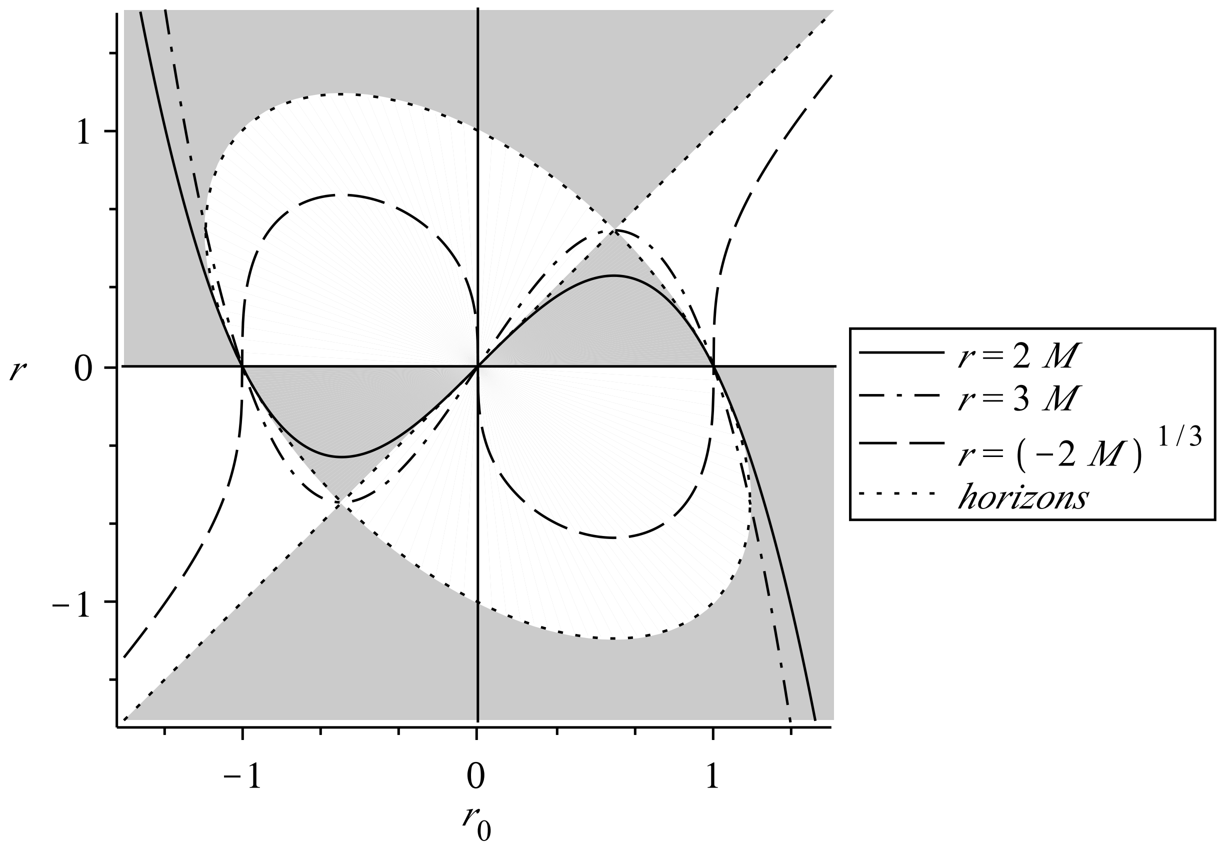

Figure 4.1: A scale-invariant picture of the SdS geometries in the (r, r_0)-plane, with various interesting values of r plotted as functions of r_0. In shaded and unshaded regions, $g_{rr} < 0$ and $g_{rr} > 0$, respectively.

restricting the ranges of either r or r_0 to positive values. However, unless $2M \equiv r_0 - {r_0}^3 = 0$, $r = 0$ is a singularity of the geometry; so it should be impossible to physically move from one side to the other. Furthermore, when $M \neq 0$, the solution is actually very different on either side of the singularity. It is therefore remarkable that when $M = 0$, and the central singularity vanishes, the metric coefficients become quadratics, indistinguishable on either side of $r = 0$, and the line-element provides the description of de Sitter space that is illustrated in Fig. 3.3.

The overall completeness of this picture alone, is suggestive of an actual significance of the entire (r, r_0)-plane; for it is clear that the conventional restriction to a physical region describing black hole geometries, say, e.g., to the interval $\{(r, r_0) \in \mathbb{R}^2 : r \geq 0, 0 \leq r_0 \leq 1/\sqrt{3}\}$, destroys that completeness.

But aesthetics alone can do no more than suggest the appropriateness of examining the entire (r, r_0)-plane in more detail; and it little disturbs *physical* intuition, to note that the entire plane could actually be allowed analytically; therefore, if we are to consider any more of it than the above interval, we must examine the physical implications of this extension. This will be important for us, as we now come to examine how the general solution describes physical mass.

4.2 Determination of Physical Mass

The reason for the problem of determining the physical SdS mass is well-known: the standard methods of calculating mass in general relativity theory require spacetime to be asymptotically flat [211]. A detailed description of these various definitions of mass, which will not be considered further here, is given in [212]. Instead, as we shall see, a good definition of mass is likely the one given implicitly in the geometries, which has profound consequences for the large-scale cosmography we will eventually use those geometries to describe.

Certainly the most interesting region of parameter space, on which we shall primarily concentrate our discussion, is that of the under-critical geometry, which describes spacetime outside a non-spinning, chargeless mass. Here, the weak field limit allows us to identify the Newtonian gravitational potential (when $\Phi \ll 1$) as

$$\Phi \approx -\frac{M}{r} - \frac{1}{6}\Lambda r^2, \tag{4.16}$$

which establishes M as roughly the spherical mass, so long as the weak field assumption is satisfied, and describes positive (negative) Λ's effect as a repulsive (attractive) force of magnitude $\frac{1}{3}\Lambda r$ [213]. This justifies our use of the term *mass parameter* in describing M. Many—e.g. [214, 215, 216, 217]—have interpreted M exactly as the SdS mass; however, there is good reason to suspect that this is not true, and more recently there have been other conjectures as to the nature of the true physical mass of the geometry (e.g., see [218], and references therein).

In fact, it can immediately be made clear, according to the description provided by Eq. (4.16) and its relation to the relativistic field, that the interpretation of M as the central mass in the geometry is inconsistent with Newtonian dynamics. For we are at liberty to use this expression for Newtonian potential to form physical intuition, because, although the gravitational field may be strong near the horizon when M is small, it is not so in general, and the relation between *physical mass* ($\equiv m$), Λ, and the radii of interest in the under-critical geometry must be consistent.

And it is indeed true that setting $m = M$ is inconsistent with the fact that a black hole's horizon radius, when $\Lambda \gtreqless 0$, is at $r \gtreqless 2M$, as well as the fact that circular photon orbits are located at $r = 3M$ for all real values of Λ (cf. Fig. 4.1; the latter point, which is proven, e.g., in [117, 118], can be checked by setting the derivative of Eq. (4.28), below, equal to zero). In order to understand why this is so, let us begin with the latter: that photon circular orbits occur at the same radius as they do in the pure Schwarzschild geometry. Then, since Eq. (4.16) tells us that Λ would act as a repulsive or attractive force, depending on its sign, we should expect the radius of photon orbits to be at $r < 3m$ if $\Lambda > 0$, or at $r > 3m$ if $\Lambda < 0$, in order to balance this additional force; i.e., according to the fact that $r = 3M$ *is* the radius of circular photon orbits, we should actually expect that $M \lesseqgtr m$, if $\Lambda \gtreqless 0$.

And the fact that the black hole horizon radius is actually at $r = R_{bh} > 2M$ ($< 2M$) for $\Lambda > 0$ (< 0), is even more troubling, if M should be thought of as the mass, and the horizon as the radius beyond which photons become 'trapped' by its gravitational field. For $\Lambda > 0$ (< 0) adds a repulsive (attractive) force to the gravitational field of the mass, which is proportional to distance, and could therefore only act to decrease (increase) the radius of such a horizon, by providing an additional nudge in the direction dictated by Λ's sign.

To clarify this point, consider the case, $\Lambda > 0$: if a black hole should trap all photons at a certain radius, which is $r = 2m$ when $\Lambda = 0$, physical intuition tells us, that the addition of a repulsive force could allow photons to escape from just inside $r = 2m$, so that the event horizon radius would *decrease* appropriately. It is therefore very troubling to that sense of intuition, that R_{bh}, which is expected to be less than $2m$, is actually greater than $2M$, *if M were the physical mass*.

Essentially the same argument has been used twice, in attempting to understand how both the radius of photon orbits and the horizon radius are altered in the presence of non-zero Λ. However, while the former should be valid, because $r = 3M$ is a well-defined spatial coordinate, an appeal to Newtonian intuition to help explain the latter point is actually false, because r is not spacelike 'beyond the horizon', regardless of the strength of the effective Newtonian force in the gravitational field. In fact, the latter argument should be no more thought of than the notion of a true Newtonian analogue to a black hole, which should have been expunged from both the scientific literature and popular expositions, following a letter from George McVittie, in 1978 [219], which provides the essential difference between the Schwarzschild and 'Laplacean' black holes: in the case of the former, according to the standard theory of black holes, light cannot escape from within $r = 2m$—it can travel in any spatial direction, but must always travel forward in time, towards $r = 0$; but in the case of the latter, light cannot escape *to infinity* from within $r = 2m$. McVittie's proof was, in fact, that a corpuscle of light emitted radially outward from a positive radius within $r = 2m$ *must* reach a finite exterior radius before it turns around.[2]

The *event horizon* is therefore a purely relativistic concept, with no Newtonian analogue; so it must be understood within a relativistic setting. In order to do that, we should consider a photon's trajectory, according to an observer, O, at some fixed radius, $r = R_O$, outside $r = R_{bh}$. If $\Lambda = 0$, the relationship between the black hole's mass and its horizon radius is well known, and, according to the discussion in § 3.2, we say that if the observer emits a photon towards $r = 0$, it will take an infinite amount of *cosmic time* to travel a coordinate distance, $R_O - 2m$. So, in order to determine whether the relation between mass and horizon radius is altered by Λ, the relevant question to ask, is whether this coordinate distance, travelled by a photon in infinite cosmic time, should be altered when $\Lambda \neq 0$.

But in de Sitter space, in which local spacetime frames describe spatial slices which expand or contract as time progresses, we know that a photon must accordingly take more or less time to traverse any given coordinate distance than it would when $\Lambda = 0$, because the corresponding physical distance would increase or decrease, respectively.

So, consider our observer outside the black hole, and allow him to vary the cosmological constant in his spacetime: $r = 2m$ remains a physical coordinate of his geometry, whether or not it is still the horizon radius. If space expands, he says the photon approaches the coordinate $r = 2m$ *more slowly* than if it were zero; but this can have no effect on the coordinate distance the photon will travel in infinite cosmic time. Similarly, if space were contracting, and the coordinate velocity of the photon were increased, it would still have to approach $r = 2m$ asymptotically.

[2]Note that Newtonian dynamics does not describe an analogue of the SdS black holes when $\Lambda \neq 0$: if $\Lambda > 0$, the repulsive force causes all corpuscles to escape to infinity, and when $\Lambda < 0$, the attractive force has the opposite effect, so that all light corpuscles would eventually be forced to turn around and fall back in.

In fact, according to the analysis in § 4.1, it should already have been quite clear that M is not likely the actual mass of the central particle, since it is not an intrinsic parameter of the geometry, but a function of the cosmological constant and the horizon radius—the parameters arising from the derivation of the metric which are physically and geometrically meaningful;—i.e., it is obvious from Eq. (4.12), that a loss of information has already occurred whenever all the integration constants are lumped together as $2M$. And it is relevant to note that the more pertinent formulation of the line-element, Eq. (4.13), does not describe the horizon radius as something that is *shifted* by an influence from Λ, but as a *triplet*, according to the cubic that formally describes the geometry when $\Lambda \neq 0$.

And these points alone should already suggest that the physical mass is equal to half the horizon radius, as the above argument also indicates;[3] then, if we set $r_0 = 2m$ in Eq. (4.12), we have,

$$2M = 2m - \frac{\Lambda}{3} 8m^3, \tag{4.17}$$

so that $\Lambda = 0 \Leftrightarrow M = m$, $2M = 0 \Leftrightarrow 2m = 0, \pm\sqrt{3/\Lambda}$, and the weak field potential, Eq. (4.16), can be rewritten,

$$\Phi \approx -\frac{m}{r} + \frac{4\Lambda}{3}\frac{m^3}{r} - \frac{1}{6}\Lambda r^2. \tag{4.18}$$

Then, there are really two possibilities for which the requirement that $\Phi \ll 1$ would hold: either the three individual terms must each be very small, or the magnitude of the additional positive term must be roughly equal to that of the two negative terms. The former turns out to be sufficiently satisfied by requiring $\Lambda \ll 1/r^2$ and $m/r \ll 1$, since $\Lambda m^3/r \ll m^3/r^3 \ll m/r$ is then already sufficiently negligible in comparison to the other two terms, so that m can be interpreted as the central mass; on the other hand, the latter case amounts to a requirement that $\frac{\Lambda}{3}r^3 \approx -2M$—roughly where the gravitational potential vanishes identically, and spacetime is locally Minkowskian (see Fig. 4.1).

Now, the problem with the interpretation of M as the physical mass manifests itself further when we attempt to use it to parametrise a continuous physical picture. For then as M increases, the two horizons at positive radii in the under-critical geometry come together, and then simply vanish, as depicted, e.g., on the interval $r_0 < -1$ in Fig. 4.1, in which M strictly increases. If we think of the traditional idea of a black hole that would be described by the under-critical geometry, in which everything that exists inside the black hole's absolute event horizon must be moving towards $r = 0$, and everything existing outside the cosmological horizon must be moving strictly outwards, then this picture implies that if M would increase past the critical value, something must happen to these separate existences, so that everything would subsequently have to move in the same direction along positive r.

But, keeping this traditional black hole picture in mind for the moment, let us now consider the description that is afforded when the mass is thought to equal precisely half the horizon radius, so that we may instead consider variations of r_0 as being equivalent to variations of the physical mass. Then, *if* no additional constraints would be imposed beyond

[3]Note that the global independence of this relation from the value of Λ was also advanced recently by Thanu Padmanabhan, through a thermodynamical investigation into the spherically symmetric solutions of Einstein's equation [220].

this requirement, the under-critical geometry would at once have three masses, for it is that triplet that is the relevant geometric parameter; i.e., whether or not this is actually correct of *mass*, it is true that the *horizon radius* is one triplet.

But what if, for example, only the horizon timelike separated from the singularity was directly related to the particle's physical mass?—so that, as the conventional interpretation would have it, the outer horizon is the cosmological horizon due to the curvature of the vacuum, perturbed by the massive presence, and the negative horizon is not physically important. Then, let us consider for the moment only the line $r = r_0$ as representing twice the value of the mass; and to begin with, consider only positive values, assuming that only the first quadrant of Fig. 4.1 has any connection with physical reality: as mass is increased, the two positive horizons come together and meet; then when mass is further increased, they cross paths, and in the intersection of the two 'oppositely directed timelike regions', r is spacelike; as mass is further increased, the 'perturbed' de Sitter horizon eventually reaches $r = 0$, then negative radii, before eventually vanishing when $r_0 > 2/\sqrt{3}$.

Conceptually, this picture is not challenging, and we might therefore like to end with this description; but in order to form it, we neglected the formal equivalence between the various parameter ranges, which requires, e.g., that an observer in the spacelike region of r 'inside' a very massive black hole, for which $1/\sqrt{3} < r_0 < 1$, cannot know whether the 'actual' mass is half the value of the observed inner or outer horizon; so our prior interpretation of a boundary separating 'inside' from 'outside' is no longer justified. Furthermore, since we should have no problem in allowing mass to increase beyond the point at which the perturbed de Sitter horizon vanishes to within $r = 0$,[4] we would also have no cause to disallow *a priori* the possibility of negative mass. Then, the logical extension requires the entire range, $-\infty < r_0 < \infty$, as well as the superposition of horizon radii, where appropriate, and that these values always equal twice that of the physical mass.

The implication that mass may also be negative is particularly interesting. In fact, negative mass is anyhow a solution of relativistic quantum mechanics, which had been seriously considered by Paul Dirac prior to the discovery of the positron (see, e.g., [221] for a description of this problem by him, from 1930, to which he proposes his 'hole' theory as a potential solution). More recently, speculation about the gravitational interaction of anti-matter has led to the proposal that although particles of anti-matter should be gravitationally attracted to each other, they might repel ordinary matter (see, e.g., [47, 48], and references therein). We shall see shortly that this is indeed true for 'negative mass' solutions, with $-1/\sqrt{27} < M < 0$, in the current theory.

But first, we should round off our inquiry up to this point by considering the radius of the outer horizon, which is decreased relative to that of the pure de Sitter geometry. It was suggested above, that the radius of the de Sitter horizon might be considered as being perturbed by the dynamical influence of a central mass; however, the effect of adding an attractive force to the centre of the de Sitter geometry, should only be to slow the escape of particles due to Λ's repulsive force, and would thus suggest that the horizon radius should increase, if anything—i.e., if the de Sitter horizon really were a spatial radius beyond which no information could ever reach the centre because of this repulsive force, then adding an attractive central force could only possibly increase that horizon radius, according to New-

[4] There is no necessity for invoking cosmic censorship in this case, as shown below.

tonian dynamics. Newtonian intuition obviously fails completely in this instance, and there is actually no apparent cause for that to be, other than the one we have just found: that the radius of the 'cosmological horizon', too, is in direct proportion to the central particle's mass.

Now, as our interpretation of physical mass has been made through an argument which has often utilised the traditional concept of a black hole, it is important to consider the validity of actually describing certain regions of r as 'timelike' and others as 'spacelike'; and, in fact, the validity of referring to the horizons as *horizons*—a term which is commonly taken to mean a one-way membrane that may be traversed, beyond which particles must be drug ever further, unless they encounter a *real* singularity. For, according to the discussion at the end of § 3.2, we can now say that there should be no actual black holes in our Universe, created through the gravitational collapse of a massive object which already fell through its horizon, leaving a vacuous timelike region in its wake—but that all such objects should instead be 'frozen stars', asymptotically approaching those limiting spatial radii, with interiors that must also be described as spatial coordinates in the Universe, since $r_{\rm Sch} < 2m$ should not be real. As well, in our discussion of Eddington's statical coordinates of de Sitter space in § 3.3, we saw explicitly that the 'timelike' interval 'beyond the horizon' is not defined; and, later on, that the coordinate singularity at $r_{\rm Edd} = \sqrt{3/\Lambda}$ in that frame, though similar, is formally different from the central particle's cosmological horizon in the Lemaître-Robertson universe, beyond which space indeed remains spacelike, so that the coordinate singularity in Eddington's system should not truly be thought to be removed in switching to the Lemaître-Robertson system.

But finally, consider the two very different physical scenarios to which the under- and over-critical geometries pertain, according to the discussion at the end of § 3.3 which led into the present analysis: Eq. (3.53) was stated as the line-element that describes isotropic motion of photons through space, relative to some point in a three-dimensional universe, as null lines; whereas Eq. (3.55) was written to describe a certain physical scenario in which the universe itself evolves isotropically from some beginning, with *cosmic* time r. Even so, these line-elements are supposed to describe events in spacetimes that actually take place on the de Sitter sphere.

Now, in the over-critical line-element, $r > 0$ is actually timelike, and this coordinate should therefore be real by design; but in the under-critical line-element, although r is spacelike on two separate intervals which are each bounded by singularities corresponding to the position of the central mass and its three-fold value, it is actually 'timelike' everywhere else. It is therefore reasonable to infer from these three points—the facts that $r_{\rm Sch} < 2m$ and $r_{\rm Edd} > \sqrt{3/\Lambda}$ do not exist, and the physical scenarios to which the two different geometries pertain,—that r is not actually real in these 'timelike' intervals, so that the under-critical line-element only provides a valid description of spacetime where r is spacelike.

Accordingly, we say that the critical SdS mass—$m = \{\pm 1/\sqrt{3}, \mp 1/2\sqrt{3}\}$, or, expressed dimensionally, $m = \{\pm 1/\sqrt{\Lambda}, \mp 1/2\sqrt{\Lambda}\}$ ($\sim 10^{53}$ kg, roughly the mass of the observable universe)—separates the geometries pertaining to two formally distinct physical scenarios, even though the global parametrisation of the line-element according to m is continuous; and that the critical mass is therefore the finite limit for such a body to have a well-defined *exterior*.

Now, the fact that the physical mass in the SdS line-element is a relativistic superposition of three values when it is below this limit, is suggestive of a quantum nature of any particle that can be described as existing in space. As I've said at the end of § 3.3, there seems to be a basic connection between this 'quantum' theory of relativistic mass and de Broglie-Bohm theory, and in § 2.4, a number of passages were quoted from Bohm and Hiley's *The Undivided Universe* [180], which is the major reference for this interpretation of quantum theory. The sentence that immediately follows the first quotation that was given there, which came from Hiley's preface to the book, is 'In the ontological theory that we present here, this wholeness is made manifest through the notion of nonlocality, a notion that is seemingly denied by relativity.'

But now we see how this problem should be answered in our dynamical theory: the 'gravitational field' due to a particle of matter is not to be thought of as the result of an action that truly warps the surrounding spacetime, thereby influencing the trajectories of surrounding particles; rather, when the causal interactions of real particles, existing in time, are described by the appropriate local spacetime line-element, it is the geometry corresponding to the perception of existence in that frame that would be perceived as such a field. The crucial point is that the pertinent spacetime geometry should thus really be only a matter of perspective, as Eddington indicated. Of course, we have seen this only in the simplest case, of an uncharged, non-spinning spherically symmetric mass, but the inference is suggested nonetheless—especially in light of the fact that our basic universe *is* (diffeomorphic to) $SU(2)$.

A formal connection between this theory and one like the de Broglie-Bohm theory, that would describe the quantum field, is beyond the scope of our current investigation. However, it is relevant to finish this note by reproducing one final passage from Bohm and Hiley's book [180], which further indicates a basic connection between their theory and the current one:

> One of the main new ideas implied by this approach [the one discussed in the final chapter] is that the geometry and the dynamics have to be in the same framework, i.e. that of the implicate order. In this way we come to a deep unity between quantum theory and geometry in which each is seen to be inherently conformable to the other. We therefore do not begin with traditional Cartesian notions of order and then try to impose the dynamics of quantum theory on this order by using the algorithm of 'quantisation'. Rather quantum theory and geometry are united from the very outset and are seen to emerge together from what may be called pre-space.

So, if an SdS particle with mass under the critical limit may be expected to have such a quantum nature, then we should say that even a galaxy cluster, with roughly zero average angular momentum and charge, and mass much less than that of the whole observable universe, could be thought of as the result of a small vacuum fluctuation. Then, we could infer that the universe we actually mean to describe—the comoving universe given by Eq. (3.44), which we've supposed to 'begin' as the 3-sphere of radius $\sqrt{3/\Lambda}$, as described in that coordinate system—might be filled with such low-mass 'particles', with *positive* masses, satisfying $0 < M < 1/\sqrt{27}$, all moving in the same direction with respect to some parallelisation of the universe, and *negative* masses, satisfying $-1/\sqrt{27} < M < 0$, moving oppositely, so that the net mass in that frame is actually zero.

On the other hand, we've found, according to the same physical hypotheses, that this universe should appear to all observers with positive mass, who measure their existences relative to the corresponding rest-frame, as an over-critical SdS universe, which must have one well-defined negative mass, $m < -1/\sqrt{\Lambda}$ ($M > 1/\sqrt{27}$), and begin from the geometrical singularity at $r = 0$; i.e., although the universe should actually have no net mass, and be spherically closed, expanding from a finite-sized beginning, such massive particles would thus perceive it as unbounded (in the coordinate t), with an overall finite negative mass, as one borne out from an initial singularity. As such, the average density of matter in finite volumes of space, large enough that small-scale deviations from the average will even out, should be an infinitesimal negative, so that the mass of the whole infinite space will have a finite negative value. Therefore, roughly equal amounts of the negative mass 'anti-matter' should still have to be observed along with the regular sort, as we expect according to the homogeneity that has been required in our cosmology, or according to standard quantum theory, if that is indeed a theory to which our results pertain.

Now, by inspection of Fig. 4.1, negative SdS masses appear to have naked singularities which are spatially separated from external observers. However, in order to determine whether this should actually be a problem, we must look at their gravitational potential, which we'll find quite revealing.

So we begin by writing the geodesic equations for both timelike and null orbits in the SdS geometries, in the equatorial plane, $\theta = \pi/2$. By substituting $2M \equiv r_0 - {r_0}^3$ into the scale-invariant SdS metric, Eq. (4.13), and writing t more explicitly as the timelike parameter, the Lagrangian for timelike geodesics with proper time, τ, may be written,

$$L = -\frac{r - 2M - r^3}{r}\left(\frac{dt}{d\tau}\right)^2 + \frac{r}{r - 2M - r^3}\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2 = -1; \tag{4.19}$$

and that of null geodesics with affine parameter, λ, as,

$$\tilde{L} = -\frac{r - 2M - r^3}{r}\left(\frac{dt}{d\lambda}\right)^2 + \frac{r}{r - 2M - r^3}\left(\frac{dr}{d\lambda}\right)^2 + r^2\left(\frac{d\phi}{d\lambda}\right)^2 = 0. \tag{4.20}$$

Because Eqs. (4.19) and (4.20) are independent of both t and ϕ, there are two conserved quantities in each case, according to the Euler-Lagrange equations, so we have,

$$\frac{r - 2M - r^3}{r}\left(\frac{dt}{d\tau}\right) = \gamma, \tag{4.21}$$

$$r^2\left(\frac{d\phi}{d\tau}\right) = J, \tag{4.22}$$

and,

$$\frac{r - 2M - r^3}{r}\left(\frac{dt}{d\lambda}\right) = \tilde{\gamma}, \tag{4.23}$$

$$r^2\left(\frac{d\phi}{d\lambda}\right) = \tilde{J}. \tag{4.24}$$

Then, after inserting Eqs. (4.21) and (4.22) into Eq. (4.19), and Eqs. (4.23) and (4.24) into Eq. (4.20), we obtain,

$$\left(\frac{dr}{d\tau}\right)^2 = \gamma^2 - \frac{r - 2M - r^3}{r}\left(1 + \frac{J^2}{r^2}\right) \tag{4.25}$$

and

$$\left(\frac{dr}{d\lambda}\right)^2 = \tilde{\gamma}^2 - \frac{r - 2M - r^3}{r}\frac{\tilde{J}^2}{r^2}. \tag{4.26}$$

which define the scale-invariant *effective potential* for a test-particle's trajectory:

$$V_{\text{eff}}(r) \equiv \frac{r - 2M - r^3}{r}\left(1 + \frac{J^2}{r^2}\right), \tag{4.27}$$

and

$$\tilde{V}_{\text{eff}}(r) \equiv \frac{r - 2M - r^3}{r}\frac{\tilde{J}^2}{r^2}, \tag{4.28}$$

as the difference between its conserved energy per unit rest-mass, or *specific energy*, and its proper radial velocity (the general relevance of using these common definitions from the analysis of particle trajectories in the pure Schwarzschild geometry, e.g., as described in [41, 222, 223], is discussed in more detail in § 4.3). But then, when $1/\sqrt{27} < M < 0$, *only* exactly radially moving photons, with $\tilde{J} = 0$, could ever reach the singularity at $r = 0$, where the effective potential is infinite, and repels all positive mass particles as well as otherwise directed photons; therefore, although the spacetime structure depicted in Fig. 4.1 suggests otherwise, the future timelike separation between a theoretical singularity with negative mass and any positive mass external test-particle, must be infinite; and furthermore, since such a mass would also repel all but precisely directed photons, it would tend to appear dark amongst surrounding positive mass galaxies (and *vice versa*). However, this is a very different form of 'dark matter' than the halos which are currently thought to influence galactic dynamics and gravitational lensing, which suggests the need for an inversion of that conception of the problem.

In essence, the picture that now presents itself to us, according to Eq. (4.27), which is plotted in Fig. 4.2 for representative values of the mass parameter, after setting $J = 0$, is that of an initially homogeneous distribution of equal amounts of matter with positive and negative mass, with a natural tendency to dissociate into clusters of each species; all presumably with $|m| \ll 1/\sqrt{\Lambda}$, so that the large-scale structure would remain homogeneous. However, to the inhabitants of any such cluster, the universe would appear to be filled only with clusters of the same form of matter, all separated by void regions of space—which would actually have been cleared of all visible matter by the dark anti-matter clustering there.

In order to describe roughly the local gravitational field surrounding a cluster which is large enough that its overall angular momentum should be negligible, consider that the mass density in its very near vicinity would be roughly equal to the value of its mass divided by the volume of the surrounding vacuum, as that would be cleared through this dissociation,—but that, as we've said above, the total density, on scales large enough that the mass distribution can be considered to be approximately homogeneous, and therefore sum to zero, must equal zero as well; therefore, the homogeneous clustering of matter surrounding any particular one, should result in an effective shell with mass roughly equal to negative that of the central

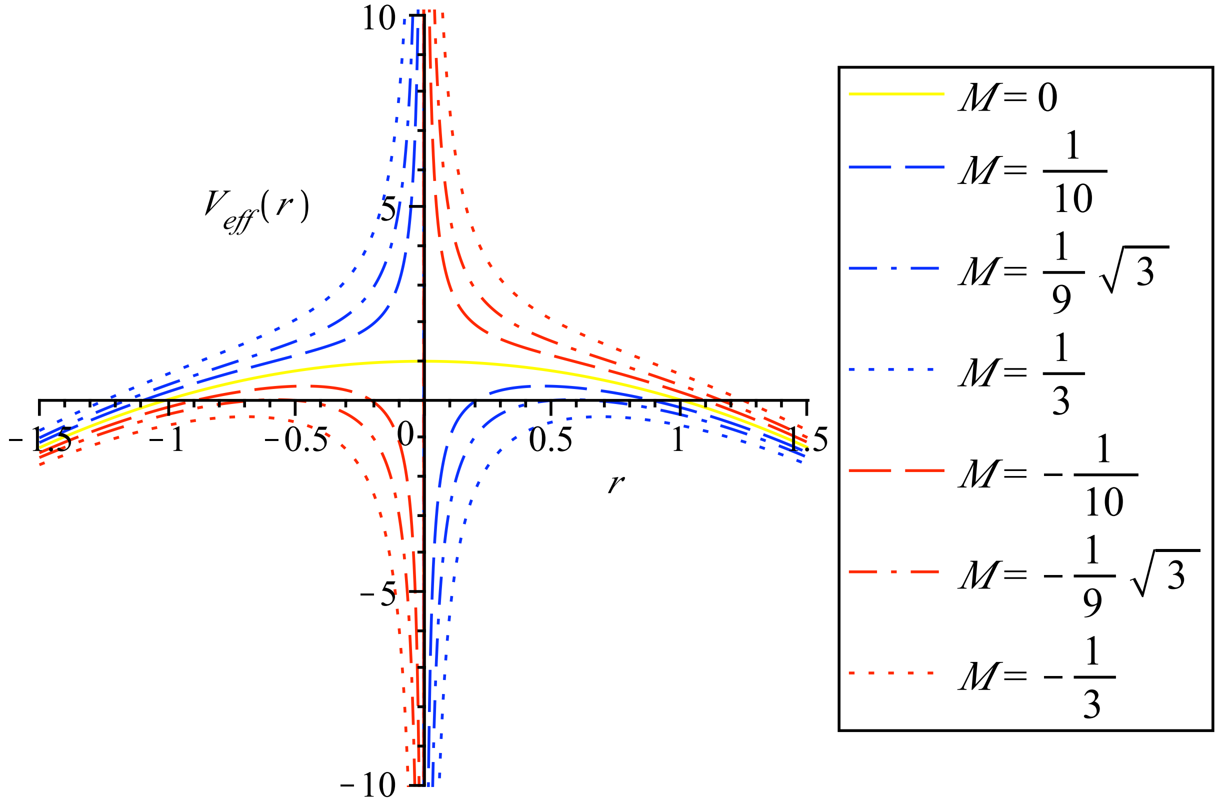

Figure 4.2: Scale-invariant effective potential curves for timelike geodesics with $J = 0$, of SdS geometries with various values of M. Representative curves are plotted for the statical de Sitter spacetime, and for under-critical, critical, and over-critical values of both positive and negative 'mass'.

cluster. Then, because the matter in that shell would actually repel the matter of which the central cluster is comprised, it should increase the effective gravitational attraction towards the centre—i.e., the existence of an effective negative mass shell surrounding a galaxy cluster, should affect its gravitational dynamics in much the same way as positive mass dark matter halos are thought to do—as the surrounding 'voids', which would actually repel visible matter, would help to bind galaxy clusters more firmly than if they existed by themselves. As such, the large-scale appearance of this universe would be similar to that of our own.

And finally, it should not be difficult to imagine, that the increase in gravitational entropy, as the two forms of matter would cluster individually, could balance the decrease in thermodynamical entropy, according to which the universe would have been formed in a hot, homogeneous state. The constancy of this entropy is in fact demanded by the SdS geometry, as it must equal one-quarter of the horizon's 'surface area' [220]; or, if the perceived mass of the universe were only slightly greater than the critical value, $|m_{univ}| \gtrsim 1/\sqrt{\Lambda}$, the corresponding constant, total entropy would be $S_{univ} \gtrsim 4\pi/\Lambda$.

4.3 Modelling the Cosmology

We are finally set to determine the physical model that would correspond to observations within the cosmic rest-frame in this cosmology; and, in order to do this, we shall begin by transforming the statical line-element of the SdS geometry into the proper coordinate system in that frame. To this end, we shall begin from the equations of motion for timelike geodesics, Eqs. (4.21), (4.22), and (4.25). Now, Eq. (4.22) defines J as a particle's specific angular momentum, which must be set to zero for all particles in the cosmic rest-frame, since t is already defined as being parallel to their motion through the comoving universal 3-sphere given by Eq. (3.44). In order to properly interpret γ, however, we require some more care.

It is helpful to momentarily restore dimensionality to the equations of motion, so that we may consider the Schwarzschild solution as a special case. With $J = 0$, these reduce to only the two equations,

$$\frac{r - 2M - \frac{\Lambda}{3}r^3}{r}\left(\frac{dt}{d\tau}\right) = \gamma, \tag{4.29}$$

$$\left(\frac{dr}{d\tau}\right)^2 = \gamma^2 - \frac{r - 2M - \frac{\Lambda}{3}r^3}{r}; \tag{4.30}$$

so that, when $\Lambda = 0$, we can immediately identify the specific energy of a particle that would remain at rest at $r = \infty$ as the conserved quantity, $\gamma = dt/d\tau = 1$. In fact, as shown in § 25.3 of [41], the term 'energy at infinity' applies generally to the term γm_0, for any particle with rest mass m_0 anywhere in the Schwarzschild geometry, so that the value $\gamma = 1$ need not be identified only with a particle that falls freely to or from rest at $r = \infty$, but actually, that its presence there need not ever be required—i.e., the term 'energy at infinity' applies for any orbit at any r in the Schwarzschild geometry, even if it cannot reach an observer at infinity who has the capability of measuring that quantity.

But the value $\gamma = 1$ does in fact have a more global interpretation in the SdS geometries, for which we find further evidence in noting that when M is set to zero instead of Λ, it corresponds to a particle that would come to rest at the one inertial coordinate in the statical de Sitter frame; i.e., if we require $dr/d\tau = 0 = d\phi/d\tau$ in Eqs. (4.21), (4.22), and (4.25), Eq. (4.25) reduces to $\gamma^2 = 1 - r^2$, so that the specific energy of an inertial partile that would remain at $r = 0$ is given by $\gamma^2 = 1$ as well. Conversely, if we set $\gamma^2 = 1$ to begin with, we find that the requirement for a particle to be spatially at rest, $dr/d\tau = 0 = d\phi/d\tau$, is satisfied only at $r = 0$. In contrast to the pure Schwarzschild case, the value of γm_0 in this geometry could therefore be referred to as a particle's 'energy at the origin'.

But in general, it should be clear that the value $\gamma^2 = 1$ really corresponds to a particle that would be at rest in any SdS geometry where its gravitational potential vanishes identically—i.e., at $r = -\sqrt[3]{2M}$, where $V_{\text{eff}} \equiv 1$. And this fact can actually be used to define all particles with *specific energy* $\gamma = 1$ as the inhabitants of a preferred frame in these geometries, since (i.) t is always timelike at this coordinate, where the metric is actually identical to Minkowski space, so it is indeed appropriate to interpret γ as the specific energy of a particle relative to that location, (ii.) it must also be that any particle with $\gamma \neq 1$, whose proper time would actually be different from t there, according to Eq. (4.21), has some constant motion relative to a particle that would come to rest there, and (iii.) according to the same point that was argued in [41] for the Schwarzschild geometry, it really doesn't make a difference to this interpretation if such a particle cannot physically come to this point, which is indeed only possible outside a negative SdS mass. Therefore, we interpret γ as a particle's 'specific energy in the cosmic rest-frame' of any SdS geometry—and, e.g., set $\gamma = 1$ in order to describe the trajectory of an inertial particle that is 'at rest' in the cosmic rest-frame of an over-critical SdS universe.

Before proceeding with the relevant coordinate transformation to the line-element that describes this frame, let us first consider an example which further illustrates the meaning of this point. As noted already in § 1.3 (cf. Eq. (1.13)), a test-particle can remain at rest at $r = M^{1/3}$, in the largest possible 'circular orbit' around an SdS mass with mass parameter M. This radius is found by solving $dV_{\text{eff}}(r)/dr = 0$ for r, after setting $J = 0$ as the lowest value of angular momentum that may be allowed for such an orbit. Now, the total specific energy of such a particle is given by $\gamma^2 = 1 - 3M^{2/3}$, according to Eq. (4.25); however, rather than interpreting this as corresponding to the particle's potential energy as it remains at rest in space in the gravitational field, we interpret it as corresponding to the difference between its specific energy in the cosmic rest-frame and a kinetic energy term due to the fact that it is not 'at rest' in *that* frame. This simply illustrates the equivalence of gravitational mass and inertial mass, that is described by general relativity theory.

We shall now determine how, in general, the SdS geometries would be perceived in the frame of an observer with $\gamma = 1$, by finding the coordinate system that describes the bundle of timelike geodesics which move through t and r all at the same rate, τ, and occupy constant positions in 'space'. The expressions for the proper velocities through t and r of these comoving geodesics can immediately be written from Eqs. (4.21) and (4.25) with $\gamma = 1$ and $J = 0$:

$$\partial_\tau t \equiv \frac{\partial t}{\partial \tau} = \frac{r}{r - 2M - r^3}, \tag{4.31}$$

$$(\partial_\tau r)^2 \equiv \left(\frac{\partial r}{\partial \tau}\right)^2 = \frac{2M + r^3}{r}. \tag{4.32}$$

The solution of Eq. (4.32) may be expressed in closed-form, by using the standard integral, $\int (u^2 + a)^{-1/2}\, du = \ln\left(u + \sqrt{u^2 + a}\right)$, after substitution of a new variable, $u^2 = 2M + r^3$. Taking the positive root of Eq. (4.32) for outgoing geodesics, we have,

$$\tau = \int_{r(0)}^{r(\tau)} \sqrt{\frac{r}{2M + r^3}}dr = \frac{2}{3}\ln\left(\sqrt{2M + r^3} + r^{3/2}\right)\Bigg|_{r(0)}^{r(\tau)}, \tag{4.33}$$

where the lower limit on the proper time has been arbitrarily set to zero. Thus, in this frame we can express the statical coordinate r of the SdS geometry as a function of each observer's proper time and a synchronous spatial coordinate, $r(0)$, which may be arbitrarily rescaled without significantly altering the description.

Then, as long as M is nonzero, a convenient set of coordinates from which to proceed results from rescaling this coordinate as,[5]

$$r(0) \equiv (2M)^{1/3} \sinh^{2/3}\left(\frac{3}{2}f(\xi)\right); \;\; M \neq 0, \tag{4.34}$$

[5] Note that this transformation is not valid when $M = 0$, which we are anyhow not interested in. An equivalent transformation in that case is found by setting $r(0) \equiv e^{f(\xi)}$, whence $r(\tau, \xi) = e^{\tau + f(\xi)}$, and Eqs. (4.39), (4.42), and (4.45), below, yield the line-element, $ds^2 = -d\tau^2 + r^2\left(d\xi^2 + d\theta^2 + \sin^2\theta d\phi^2\right)$.

from which we find, after some rearranging,[6]

$$r(\tau,\xi) = (2M)^{1/3}\sinh^{2/3}\left(\frac{3}{2}[\tau + f(\xi)]\right), \tag{4.35}$$

which immediately shows the usefulness of expressing the arbitrary coordinate ξ in the form written in Eq. (4.34), because it allows Eq. (4.33) to be solved explicitly for $r(\tau,\xi)$. Therefore, without loss of generality, we define a new coordinate $\chi \equiv f(\xi)$, so that $-\tau \leq \chi \leq \infty$. This has no effect on the eventual form of the metric we will derive, due to the chain rule, but allows for a neater calculation. As such, we immediately have the useful result,

$$\partial_\chi r \equiv \frac{\partial r}{\partial \chi} = \partial_\tau r = \sqrt{\frac{2M + r^3}{r}}. \tag{4.36}$$

The transformation, $t(\tau,\chi)$, may then be calculated from

$$\frac{dt}{dr} = \frac{\partial_\tau t}{\partial_\tau r} = \frac{r}{r - 2M - r^3}\sqrt{\frac{r}{2M + r^3}}; \tag{4.37}$$

furthermore, to solve for $t(\tau,\chi)$, we can gauge the lower limits of the integrals over t and r, at $\tau = 0$, by requiring that their difference, defined by

$$t(\tau,\chi) = \int^{r(\tau,\chi)} \frac{r}{r - 2M - r^3}\sqrt{\frac{r}{2M + r^3}}\,dr - F(\chi), \tag{4.38}$$

sets

$$g_{\chi\tau} = g_{tt}\partial_\tau t\partial_\chi t + g_{rr}\partial_\tau r\partial_\chi r = 0. \tag{4.39}$$

This calculation is straightforward:

$$\begin{aligned} 0 &= -\frac{r}{r - 2M - r^3} + \partial_\chi F(\chi) + \frac{r}{r - 2M - r^3}\frac{2M + r^3}{r} && (4.40)\\ &= \partial_\chi F(\chi) - 1, && (4.41)\end{aligned}$$

so that $F(\chi) = \chi$.

Now, it is a simple matter to work out the remaining metric components as follows: our chosen reference frame immediately requires

$$g_{\tau\tau} = g_{tt}(\partial_\tau t)^2 + g_{rr}(\partial_\tau r)^2 = -1, \tag{4.42}$$

[6]Note that the two identities, $e^x = \sinh(x) + \cosh(x)$ and $\mathrm{arsinh}(x) = \ln\left(x + \sqrt{x^2+1}\right)$, are useful here. Eq. (4.35), along with our eventual line-element, Eq. (4.46), was originally found by Lemaître [111], although his solution to Eq. (4.32) (with dimensionality restored),

$$r = (6M/\Lambda)^{1/3}\sinh^{2/3}\left[3\sqrt{\Lambda}(t - t_0)/2\right],$$

is too large in its argument by a factor of $\sqrt{3}$.

according to the Lagrangian, Eq. (4.19), with $\partial\phi/\partial\tau = 0$; and by direct calculation, we find

$$\begin{aligned} g_{\chi\chi} &= g_{tt}(\partial_\chi t)^2 + g_{rr}(\partial_\chi r)^2 && (4.43)\\ &= -\frac{r-2M-r^3}{r}\left(\frac{2M+r^3}{r-2M-r^3}\right)^2 + \frac{r}{r-2M-r^3}\frac{2M+r^3}{r} && (4.44)\\ &= (\partial_\chi r)^2. && (4.45)\end{aligned}$$

But this result is independent of the definition of χ, as explained above; for if we had followed through with the more general coordinate ξ, we would have found the metric to transform as $g_{\xi\xi} = g_{\chi\chi}(d\chi/d\xi)^2 = (\partial_\xi r)^2$, the other components remaining the same. Therefore, the SdS metric in the proper frame of an observer who is cosmically 'at rest', in which the spatial coordinates are required, according to an appropriate definition of $F(\chi)$, to be orthogonal to τ,[7] can generally be written,

$$ds^2 = -d\tau^2 + (\partial_\chi r)^2\, d\chi^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right). \tag{4.46}$$

Thus, we have proven Lemaître's result from 1949 [111],—that spaces with $d\tau = 0$ are Euclidean, with line-element,

$$d\sigma^2 = dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right). \tag{4.47}$$

Lemaître concluded from his calculation, that the 'geometry is Euclidean on the expanding set of particles which are ejected from the point singularity at the origin', which is represented in this frame by the line $\chi = -\tau$; but he did not proceed to develop a cosmology directly from this result. Instead, for the universe as a whole, he cast the geometry aside, writing that, in the case of such a singularity, 'the particles continuously ejected have no physical reality; they are just mathematical devices introduced in order to define the coordinates and the corresponding partition of space and time', and proceeded with the development of a FLRW cosmological model, eventually applying this result in a statement, that 'in order that spherical symmetry should exist around any point, we must particularize the results found for spherical symmetry, so that the three spatial coordinates θ, ϕ, and χ would appear in the expression $d\sigma$ only'. This, he accomplished by redefining r as a separable function of the coordinates χ and τ, which is impossible in the SdS geometry, according to Eq. (4.35), since, aside from the trivial case (x or $y \equiv 0$), the hyperbolic sine function cannot be written $\sinh(x+y) = A(x)B(y)$. As such, his new $d\sigma$—the RW spatial line element—would bear no relation to the original geometry from which it had been calculated.

Essentially what he did was to interpret the synchronous set of events in this spacetime as the comoving universe—which is false, as one should understand according to the discussion of Chapter 3; for, by the definition that was finally given in Eq. (3.55), the dynamic universe described by the over-critical SdS spacetime should actually be comoving along r.

It is interesting, and clarifies the physical meaning of this result, to recheck our calculation by beginning with the expression we've found for the cosmic radius, Eq. (4.35). According to this expression, spaces of a given radius must have the same value, $\bar{\tau} \equiv \tau + \chi$; and the set of comoving inertial observers described by Eq. (4.46), who each remain at constant χ, and

[7] Note that θ and ϕ were already orthogonal to τ.

whose clocks all tick at the same rate τ, must coexist dynamically on the causal hypersurface $\tau = \bar{\tau} - \chi$, rotated an angle $\pi/4$ from the τ-axis, in the (τ, χ)-plane; thus, $\bar{\tau}$ is the cosmic time of the bundle of geodesics defining the cosmic rest frame. And this fact may be used to rewrite the statical and proper line-elements for the SdS geometry, using notation which is meaningful to observers who are cosmically 'at rest', according to the total derivative of r with respect to the cosmic time,

$$\frac{dr}{d\bar{\tau}} = (2M)^{1/3} \frac{\cosh(3\bar{\tau}/2)}{\sinh^{1/3}(3\bar{\tau}/2)} = \sqrt{\frac{2M + r^3}{r}}, \tag{4.48}$$

after noting, from Eq. (4.38), the corresponding transformation of the coordinate t,

$$\begin{aligned} dt &= \frac{r}{r - 2M - r^3} \sqrt{\frac{r}{2M + r^3}} dr - d\chi, \\ &= -\left[\left(\frac{dr}{d\bar{\tau}}\right)^2 - 1\right]^{-1} (d\tau + d\chi) - d\chi. \end{aligned} \tag{4.49}$$

These may be substituted directly into the statical SdS metric, Eq. (4.13). Then the statical and proper line-elements are, respectively,

$$ds^2 = -\left[\left(\frac{dr}{d\bar{\tau}}\right)^2 - 1\right]^{-1} dr^2 + \left[\left(\frac{dr}{d\bar{\tau}}\right)^2 - 1\right] dt^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{4.50}$$

$$= -d\tau^2 + \left(\frac{dr}{d\bar{\tau}}\right)^2 d\chi^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{4.51}$$

from which it is obvious that the critical condition for $r > 0$ to be timelike, $M > 1/\sqrt{27}$, must be equivalent to the requirement, $(dr/d\bar{\tau})^2 > 1$. This may be confirmed using Eq. (4.48), from which we find that $dr/d\bar{\tau}$ has a minimum at the inflection point, $3\bar{\tau}_{accel} = \ln\left(2 + \sqrt{3}\right) = 2\,\mathrm{arsinh}\left(1/\sqrt{2}\right)$, of $r(\bar{\tau})$, at which particles begin to accelerate away from $r = 0$. Substituting this value back into Eq. (4.48), we find,

$$\left.\frac{dr}{d\bar{\tau}}\right|_{\sinh(3\bar{\tau}_{accel}/2)=1/\sqrt{2}} = (2M)^{1/3} \frac{\sqrt{3/2}}{2^{-1/6}} = \sqrt{3}M^{1/3}; \tag{4.52}$$

so that, indeed, $(dr/d\bar{\tau})^2 > 1$ for all $\bar{\tau} \Leftrightarrow M > 1/\sqrt{27}$.

What we would now like to understand, is how such a universe would appear to an observer in the cosmic rest-frame, who observes light which has causally propagated *through* the void *within* the non-synchronous universe; so, in order to illustrate this situation, various null-lines of constant θ and ϕ,

$$\frac{d\tau}{d\chi} = \pm \frac{dr}{d\bar{\tau}}, \tag{4.53}$$

have been plotted in Fig. 4.3, in the (τ, χ)-plane of the $M = 1/2$ ($> 1/\sqrt{27}$) SdS geometry, alongside the corresponding section of the de Sitter sphere, from Fig. 3.5.

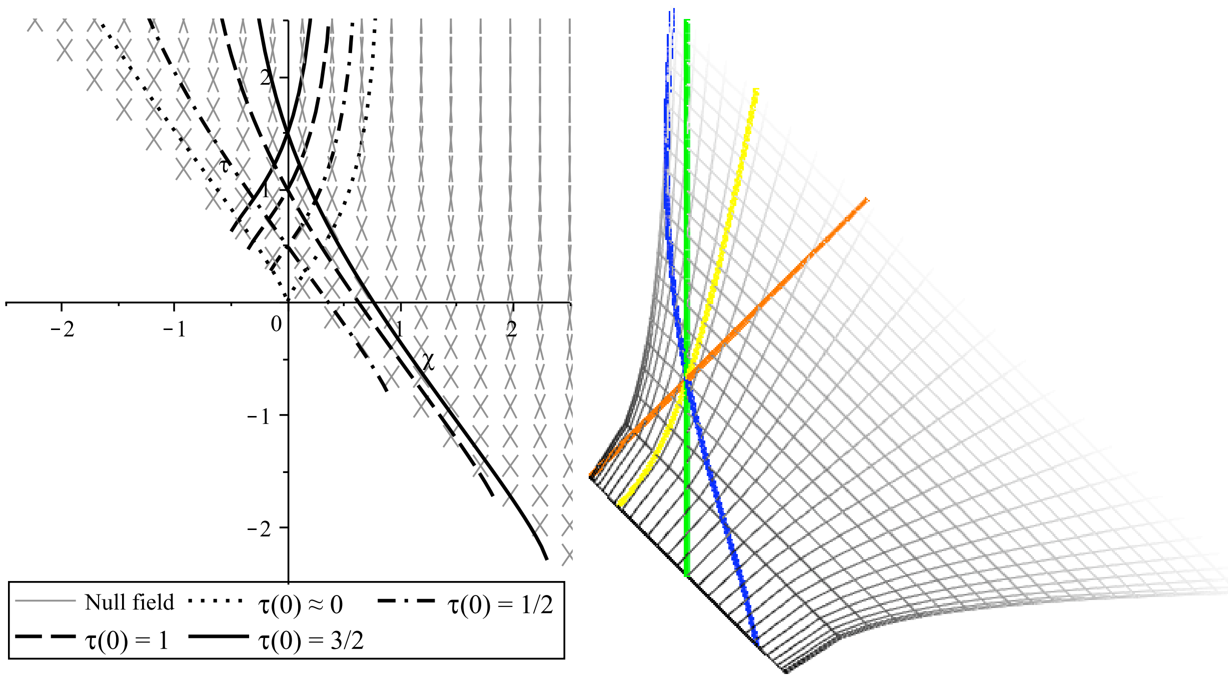

Figure 4.3: Null geodesics in the (τ, χ)-plane of the SdS geometry with $M = 1/2 > 1/\sqrt{27}$ (left), plotted alongside the corresponding section of the de Sitter sphere (right; see Fig. 3.5). A universe must evolve along one increasing radius, with cosmic time $\bar{\tau} = \tau + \chi$, comoving at a forty-five degree angle from the τ-axis. By design, particles at-rest in the cosmic rest-frame follow flat null-lines through de Sitter space, with 'simultaneous space' at any particular value of proper time described by the oppositely-directed null-line passing through that event.

With this diagram, it can be easily pictured how light approaching an observer from positive or negative χ originates from prior $\bar{\tau}$, and is emitted, in either direction, towards increasing $\bar{\tau}$, even though, at any τ, every observer perceives simultaneous space as the flat slice of spacetime ranging from $\bar{\tau} = 0$ to $\bar{\tau} = \infty$, with the 'big bang' singularity at $\chi = -\tau$ (cf. Lemaître's interpretation of particles continuously ejected from the point singularity at the origin). The associated graph of de Sitter space, in which one of the flat null-lines has been oriented vertically, so that it would correspond to the τ-axis of the particle at $\chi = 0$, shows that the τ- and χ-coordinates should be the collections of oppositely directed (flat) null-lines on the de Sitter sphere, which run parallel to each other in the coordinate system of Eq. (4.46)—i.e., the coordinate system we use to describe spacetime in the cosmic rest-frame of the 'over-critical SdS universe' takes the flat null-lines of the de Sitter sphere and lays them out orthogonally on a plane;—and the association, of null-lines in that coordinate system, to the 'photons' defined in our theory (which are represented in Fig. 4.3 by blue and yellow world-lines) is likewise apparent. As well, it is clear that massive particles should evolve between those null-lines in either picture, and that the world-lines of particles with 'negative mass'—i.e., those which move in the opposite direction through the comoving universal 3-sphere—should indeed be spacelike, as we effectively defined from the outset (i.e., at the end of § 3.3). Thus, e.g., the world-line of a particle that follows an oppositely directed null-line on the de Sitter sphere, such as the centres of mass of 'dark anti-matter clusters' postulated to exist in § 4.2, would be described as occurring all at an instant in the (τ, χ)-frame.

But this ideal spacetime description, according to which some particles or massive clusters would run their entire course of existence in an instant, and others could even be said to move backwards in time, cannot actually be an accurate description of the cosmological evolution that would be observed, since the comoving universe does not actually exist synchronously with any such observer, and when cosmological measurements happen to be made, the observer must still see the real causal evolution that has taken place in *cosmic* time.

So we must now put together the pieces of our description, in order to say precisely how physical observables should be modelled in this frame. To begin with, we note that the cosmological redshift of light emitted from anywhere in the homogeneous universe at some earlier cosmic time, $\bar{\tau}_e$, and observed presently, at $\bar{\tau}_0$, must be given—either according to the real physical expansion of the universe, or else the perceived difference in effective potential—by

$$1 + z = \frac{r(\bar{\tau}_0)}{r(\bar{\tau}_e)}. \tag{4.54}$$

Also, because cosmic rest-frame observers travel along null-lines of the de Sitter sphere, all propagating in the same direction and fanning out as cosmic time increases, the perceived expansion based on a homogeneous distribution of galaxies travelling along those fundamental worldlines, must still be isotropic from the relative perspectives of such observers. Finally, we have the two relevant[8] frame-specific aspects of perception: that the simultaneous slices of spacetime in the observer's past light-cone are flat, and that r is given as a specific function of the proper time of any such observer, so that the line-element that would be appropriate for modelling physical observables within this cosmology must be the one corresponding to a flat RW expansion, with $K = 0$ in Eq. (1.12), and,

$$a(\bar{\tau}) \equiv r(\bar{\tau}) = (2M)^{1/3} \sinh^{2/3}(3\bar{\tau}/2)\,; \tag{4.55}$$

or, restoring units ($a \mapsto \sqrt{\Lambda/3}a' = a$),

$$a(\bar{\tau}) = \left(\frac{6M}{\Lambda}\right)^{1/3} \sinh^{2/3}\left(\frac{3}{2}\sqrt{\frac{\Lambda}{3}}\bar{\tau}\right). \tag{4.56}$$

It is not difficult to derive Eq. (4.56) as the solution to Friedman's equations, Eqs. (1.9) and (1.10), for a flat ΛCDM universe with $p = K = 0$, by direct integration. However, this is unnecessary because the same can be proven simply by showing that $\dot{a}/a$, as calculated

[8] Actually, the flatness of synchronous slices in the fundamental rest-frame is probably not relevant at all, since the curvature of RW space only enters into cosmological observations through the scale-factor, and we anyhow have the result that $r(\bar{\tau})$ scales precisely as the *flat* ΛCDM scale-factor, in a universe that essentially *is*, and must also *appear* isotropic to all observers in the fundamental rest-frame; i.e., because through standard FLRW theory, all we really measure through cosmological observations is the form of a as a function of t, which must satisfy Eqs. (1.9), (1.10), and basic assumptions about the large-scale perfect fluidity of the universe; and we have already derived such a scale-factor in a very different way, as a specific function which is equivalent to the very specific such model in which $K = 0$, $p = 0$, and $\Lambda \neq 0$.

directly from Eq. (4.56), is equal to the flat ΛCDM *Hubble parameter*, since the requirement that $a(\bar{\tau} = 0) = 0$ is already satisfied. In doing so, we'll find an interesting connection to the SdS mass parameter, M.

To begin as usual, note that compatibility of Friedman's equations requires

$$\dot{\rho} = -3(\rho + p)\frac{\dot{a}}{a}, \tag{4.57}$$

which may be used in place of Eq. (1.10), so that a FLRW universe containing only nonrelativistic matter, with $p \ll \rho \equiv \rho_m$, satisfies

$$\rho_m a^3 = \rho_{m,0} {a_0}^3 = const., \tag{4.58}$$

where the subscript 0 denotes evaluation at the present time, $\bar{\tau}_0$. Then, with $K = 0$ and general Λ, Eq. (1.9) may be rewritten,

$$H \equiv \frac{\dot{a}}{a} = \sqrt{\frac{\Lambda}{3} + \frac{\kappa \rho_m a^3}{3a^3}}, \tag{4.59}$$

where H is the Hubble parameter. This is the flat ΛCDM Friedman equation, expressing $\dot{a}$ as a function of a only, which has been quite precisely constrained through observations of the CMB, SNe Ia, and other phenomena, this past decade.

But, noting that $\coth(\operatorname{arsinh}(x)) = \sqrt{1 + x^{-2}}$, this same result follows immediately when we write the Hubble parameter according to Eq. (4.56); i.e., because then

$$\begin{aligned} \frac{\dot{a}}{a} &= \sqrt{\frac{\Lambda}{3}} \coth\left(\frac{3}{2}\sqrt{\frac{\Lambda}{3}}\bar{\tau}\right), \\ &= \sqrt{\frac{\Lambda}{3} + \frac{2M}{a^3}}, \end{aligned} \tag{4.60}$$

so that, with $\kappa \equiv 8\pi$, we can immediately identify

$$M = \frac{4\pi}{3} a^3 \rho_m \tag{4.61}$$

as the apparent constant mass per unit comoving volume in the perceived flat ΛCDM universe. This quantity cannot be determined through redshift measurements though, because those correspond to the ratio, a_0/a_e, between the present scale-factor and its value at the epoch which is presently being observed, and M vanishes from that quantity. More explicitly, the luminosity distance of an object with redshift z may be calculated from ($u \equiv a_0/a$)

$$d_L = (1 + z) \int_1^{1+z} \frac{du}{H(u)} = (1 + z) \int_1^{1+z} \frac{du}{\sqrt{\frac{\Lambda}{3} + \left(H_0^2 - \frac{\Lambda}{3}\right) u^3}}, \tag{4.62}$$

so that Λ and the *Hubble constant*, H_0, can, e.g., be determined by modelling the apparent magnitude (m_{mag}) distribution of a sample of *standard candles* (all with the same absolute magnitude, M_{mag}, such as SNe Ia) according to the distance modulus equation,

$$m_{\text{mag}} - M_{\text{mag}} = 5 \log_{10}\left(\frac{d_L}{10 \text{ pc}}\right). \tag{4.63}$$

But then, according to Eqs. (4.60) and (4.61), only values such as $\bar{\tau}_0$, $M/{a_0}^3$, and $\rho_{m,0}$ can be calculated subsequently. Even so, we can say (for what it's worth) that M has to be greater than the limit set by the over-critical condition, $\sqrt{\Lambda/3}M > 1/\sqrt{27}$; or, using $\Lambda \sim 10^{-52}$ m^{-2}, $M > 1/(3\sqrt{\Lambda}) \sim 5 \times 10^{52}$ kg.

4.4 Cosmic Black Holes and an Omniverse

Let us consider finally, how such a universe as the one that has now been described should be created through the eventual gravitational collapse of a system of galaxies that existed in a prior one. There is in fact little that needs to be said in order to complete this picture, since the physical scenario that is left for us to state should already be obvious to one who has followed the prior theoretical development to this point. So, while keeping in mind some of the main aspects of the theory—viz., the fact that the ideal evolution of our universe is given by the SdS line-element, which similarly describes a spacetime that is locally dominated by a spherical mass, and that the universe should fully run its course only on the 'expanding half' of de Sitter space—we shall revisit the problem of gravitational collapse.

According to our theory, local expansion dynamics should really be occurring in roughly the manner examined by Lemaître [76, 77, 78, 111, 112, 113] (or, e.g., de Sitter [88], Eddington [2], or Pirani [116]), which was briefly discussed in § 1.3. Therefore, roughly speaking, as the Universe evolves, every galaxy or cluster of galaxies with no net angular momentum, which is denser than the limit set by Eq. (1.13), will remain gravitationally bound in the face of cosmic expansion that is driven by Λ. According to the standard theory of galactic dynamics (e.g., see [224, 107]), every such localised pocket of matter shall evolve, through the interactions of its constituent parts, into a single giant elliptical galaxy, again with roughly no net angular momentum or charge; which, through subsequent relaxing processes, shall, according to the spacetime dynamics of gravitational collapse described above, inevitably collapse to a critically dense SdS configuration, asymptotically falling towards its 'event horizon' as everything else in the Universe recedes exponentially.

According to this picture, even the collapse of a single star is not to be thought of as having a very close connection to the final state—for although the collapsing star will appear to 'freeze' at its Schwarzschild radius and remain there for a very long time, it will not remain in that solitary state forever, but shall eventually come in close contact with other bodies, and dynamically interact with them in many possible ways; i.e., although a particle on the star's surface could be thought to reach the star's Schwarzschild radius in a small interval of proper time, it cannot reach any such horizon before every eventual interaction of the frozen star has taken place and the end result of all those interactions has come to the final one just described, so that the horizon it will come to will be very different from the one it initially began falling towards.

Now we must consider the final description of this collapse. The possibilities that have anything to do with the traditional conception, in which the central regions should cross an inner horizon first, and the inner timelike interval of r should grow until all of the collapsing material has entered the black hole, are completely denied by this causally coherent description. For in fact, even if we attempt to reconcile this idea with the fact that interior particles cannot fall through inner horizons before the outer particles do, by suggesting that every particle should reach a limiting inner radius ‘at once’, so that there is a sort of ‘instant’[9] at which all particles that comprised the former ‘frozen star’ should now be distributed *throughout* the interval of now-timelike r, we still have a violation of causal coherence, as these particles should no longer inhabit one universe.

In fact, the possibility that remains, is that the particles, which all inhabited the same universe with local measure of cosmic time described by t at $r = \sqrt[3]{-2M}$ before the horizon, must *universally* collapse, with subsequent cosmic evolution occurring radially.

The picture that this brings to mind is an unfortunate three-dimensional one, involving a 2-sphere that is somehow created in space just prior to collapse, which subsequently contracts as it evolves towards $r = 0$—and that is incorrect. But rather than trying to talk about all of the reasons that this should be incorrect, according to the four-dimensional theory of relativistic dynamics, without the clarity that would be afforded by an analytical solution, it seems best to simply end this discussion with a conjecture: that the product of this collapse is a 3-sphere in de Sitter space, similar to our universe, which may also be described globally by the *over*-critical SdS line-element, with velocities randomly oriented in all directions in that collapsing 3-sphere, so that our previous assumptions about the fundamental cosmic rest-frames of positive and negative mass particles should similarly apply; i.e., so that, from the collapse of a cluster with positive *or* negative *under*-critical SdS mass, which should reach its horizon at infinite cosmic time, a collapsing *over*-critical SdS universe should emerge with a perfectly random distribution of particle velocities. This universe should then collapse to a finite radius in de Sitter space, or else all the way to the ‘singularity’ at $r = 0$ in the over-critical SdS geometry pertaining to the fundamental rest-frame of massive particles, before subsequently expanding in the manner that has been described.

Such an origin for our universe would finally answer the problem of the cause of cosmic time, that needed to be left unanswered at the end of § 3.3; for although the Lorentzian signature of de Sitter space and the homogeneity of the actual universe described by Eq. (3.44) had indicated that this was a good possibility to explore, the description that is offered by the mathematics still does not indicate what would cause the universe to be this moving 3-sphere. However, if such a universe could thus be described as the direct result of a process which has been confirmed to take place in our Universe—viz., of the inevitable final stage of gravitational collapse of massive bodies,—then we could indeed attribute this as a prior natural cause of the cosmic geometry and time prescribed in our theory, which corresponds well with cosmological observations.

Then we could infer the existence of a great number of universes similar to our own—each one the product of gravitational collapse in a progenitor universe that became a multiverse when its infinite course had ended. The logical extrapolation of this is to an omniverse: an infinite pattern of such past and future universes. Then, as the process of beginning in the omniverse should not be singular, it is reasonable to suggest that each universe should retain some information from its progenitor cluster, as similarly considered elsewhere [80, 41, 225].

[9]According to the language that was developed in § 3.1, this would be an instant in the five-dimensional ideality (if such could be given when causal coherence is not assumed *a priori*) of the four-dimensions of reality.

Chapter 5

Conclusions

> To explain the phenomena in the world of our experience, to answer the question "why?" rather than only the question "what?", is one of the foremost objectives of all rational inquiry; and especially, scientific research in its various branches strives to go beyond a mere description of its subject matter by providing an explanation of the phenomena it investigates.
>
> —Carl Hempel and Paul Oppenheim, 'Studies in the Logic of Explanation'

Our standard cosmological model fits very well with the empirical data, and the values of its parameters have been able to be constrained quite accurately due to innovations in cosmological observation. The problem with this model is that it pertains to a theory—viz. FLRW theory, which describes the dynamical evolution of space in the RW line-element according to Einstein's relativity theory—that, for all its robustness in providing a generally applicable empirical model, when it was finally subjected to data that were precise enough to allow for meaningful discrimination in the actual values of its parameters, presented us with a picture of the Universe that has little to do with the rest of our knowledge, so that it is very different from any expectations that could be reasonably based on terrestrial theory.

For, now, together with the problems of the Universe's unnatural appearance—i.e., that it appears to have little or no large-scale curvature, which basic FLRW theory describes as a statistical impossibility, and the fact that it appears to be isotropic on large scales, with local structures everywhere appearing similar even though they should not have been in causal contact prior to their formation,—we find ourselves fully confident that this unnatural Universe should also contain some form of dark matter, the presence of which has also been observed in local dynamics, along with some form of dark energy that is now driving the apparent acceleration of the cosmic expansion rate; and that these two exotic components currently make up $\sim 96\%$ of the energy-content in the Universe. As well, we have the long-standing problem of the vacuum expectation value, which is a hundred orders of magnitude greater than the cosmological constant that we do observe. Thus, e.g., it has been hypothesised that the Universe should contain some sort of quintessence field that would have driven an inflationary epoch early on, but quickly relaxed to an asymptotic state that would resemble a cosmological constant.

But, for all of these problems, the fact remains that the empirical model is good and well-constrained—so the issue is not one of accurately accounting for the data, but of reasonably explaining them.

Indeed, the greatest challenge for the experimentalists has not been to *describe* the data, but to objectively account for the number of complexities that might be imposed upon the theory—and they've consistently found that the best model is the simplest one, with only two parameters, rather than any more complex variant.

The most common epistemological method that has been used to approach the cosmological problem is the one adapted from particle physics; viz. the 'Shut up and calculate!' approach. However, the problem with this method is that it has no *explanatory* power. In quantum theory, it *has* proven to have great *predictive* power, which some might even argue is better still; but even David Mermin, who likely coined the now-famous phrase, has since admonished it as being 'snide and mindlessly dismissive', among other things [226]. Even so, it is without a doubt the most common practice amongst theoretical physicists today to employ this experimentalist-style method: to begin by modifying the basic theoretical apparatus or its input in some way—shut up and calculate—and see what new result comes from that, as it may be most readily interpreted within the context of a prior world-view.

The issue, though, is that the type of problem to which this method best applies is not the one that we've now got in cosmology. We've got reliable data, along with a model that beautifully accounts for them, according to a good idea of what is really going on. We've even found, parametrically speaking, that this is really about the simplest possibility that could have been: the empirical data indicate that the Universe is flat and isotropic, appearing the same even in regions that could not previously have been in causal contact (unless something like an inflationary epoch occurred), and that it has expanded, everywhere in absolute (CMB rest-frame) time, precisely as a hyperbolic sine function. The only real problem has been the lack of a good idea about how to explain these results within the context of the corresponding theory—but then one commonly takes to experimental theoretics, as a means of possibly predicting a small perturbation that would give rise to the apparent way of things.

But there is another, far more direct epistemological approach to the realistic explanation of unexpected empirical results; viz., to search for an objective re-interpretation of the theory's fundamental basis through rational speculation. The beauty of re-interpretation is that one does not even have to go to calculation, at least not to start with: the mathematical description is supposed to remain equivalent, while only the theory's meaning should change. Then, when all aspects have been rationally taken into account, one sees gross paradoxes and other significant problems simply fade away.

An example is the re-interpretation of special relativity theory that was presented in § 3.1. For it may be said that Einstein did interpret the relativity of simultaneity as a noumenon: he thought that it would be most objective to say that the events that occur simultaneously relative to each observer should have *really* occurred simultaneously from each individual's perspective, in their own proper frame. Then, the most straightforward interpretation that is consistent with the mathematics, is the block spacetime interpretation that he eventually believed in (§ 2.7). However, there are other options; e.g., that a noumenal present, perhaps defined as that which occurs simultaneously for each observer individually, may exist without having an objective representation; or that each observer's relative reality might be their past light cone, rather than their present simultaneous hyperplane.

But such interpretations implicitly operate from the same interpretation of the relativity of simultaneity that was originally made by Einstein; viz. that two events on an observer's past light cone that are described in their proper frame as having occurred simultaneously should be described by them as having *really* occurred simultaneously. In contrast, I have argued for a reinterpretation of special relativity theory that would describe the relativity of simultaneity as a phenomenon within a three-dimensional universe evolving along a four-dimensional manifold with Lorentzian metrical structure, in which photons travel along the invariant null lines. In this interpretation, the noumena that occur with true simultaneity will only be described as simultaneous phenomena by observers who remain absolutely at rest. In § 3.1, the situation of pens tracing out lines on a barograph sheet was used to illustrate how a particular foliation of special relativistic spacetime might be objectively described in this scenario.

With the concept of an absolute present reconciled in this objectively well-defined way (meaning that the physical description of an objective reality is equivalent in any system of cooridinates), not only is the problem of the block universe implication found to be resolved, but, since the actual events of the past and future that we commonly think of should then clearly be described as events of an *ideal* spacetime formed of noumena within a real three-dimensional universe as it really evolves, the paradoxes of time travel are trivially resolved as well.

Furthermore, if the cosmic time of standard cosmology were interpreted as such an absolute time, as it may be, with a corresponding absolute rest-frame that anyhow finds empirical support in the CMB, then the same should apply. Therefore, because such paradoxes have still been thought to exist, this re-interpretation is perhaps most significant in the way that it provides for a clear distinction to be made between the real time of the universe that presently evolves, and the ideal timeline of the four-dimensional continuum of events, in the objective covariant physical description that applies as well to both.

But there remains something unsettling about the inertial and causal structures that need to be assumed in special relativity theory, which seems to be made more manifest in the proposed scenario in which the absolute time adds further structure to the spacetime manifold. A resolution to this problem was proposed in § 3.3, where it was found that de Sitter space alone is the maximally symmetric real Riemannian manifold which has intrinsic Lorentzian signature; thus, it was recognised that if the Einstein equation for Physical Reality did have a strictly positive cosmological constant, an appropriate causal structure would be inherent in spacetime. In particular, with an appropriate definition of absolute time to describe an objective present, it then became possible to formulate a special relativistic scenario similar to the one given in the case of flat space, but which works only from philosophically well-motivated principles.

It is therefore remarkable to note, that although relativity theory *can* be formulated without an absolute time or a cosmological constant, the empirical evidence from the CMB and the success of the standard cosmological model, which suggest that there is in Reality an Absolute Time, and the evidence from the measured cosmic expansion rate, suggesting that the cosmological constant actually should be positive, really do support all of this speculation; and so, prior reasons have been found to alter our original expectations to ones which are entirely consistent with the empirical facts.

The remainder of the thesis progressed from here in roughly the same vein, to the formulation of a cosmological theory which exactly predicts the empirical observations that are made. And the key difference between this theory and the special theory of relativity that is formulated on a Minkowski background, is that although the causal structures of

Minkowski space and de Sitter space are similar, their inertial structures are in fact very different. For, when the simple definition of absolute time has been made according to the induced causal structure of a de Sitter sphere, it follows that the paths of all particles with (non-trivial) uniform coordinate velocity in the absolute rest-frame—i.e., with non-trivial *absolute motion*—are *not* geodesics.

This led to the hypothesis, that the de Sitter sphere, with its intrinsic causal and inertial structure, might actually be an *absolute space* through which the universe evolves as a 3-sphere that contracts and expands in absolute time, and that inertia might be defined as a particle's tendency to remain in uniform absolute motion, so that physical mass might then be related to the fact that the worldlines of inertial particles are typically not geodesics—as opposed to the theory that gravitational mass should truly warp space so that inertial particles would follow geodesics of the resulting curved spacetime.

This hypothesis was supposed to be supported in the following way: by recognising that in the neighbourhood of such an intertial particle it should be possible to algebraically construct a line-element that would describe a relatively symmetrical local spacetime geometry in which invariant null lines would 'forever' propagate in the 'radial' direction at the same coordinate velocity, and by subsequently solving for the components of the metric tensor in that coordinate system according to the requirement that it has to satisfy the vacuum Einstein equation, the local form of the statical SdS solution was recovered, and it was found that this indeed has one free parameter which should correspond directly to the mass so described. In particular, it so happens that according to this interpretation of the local SdS geometry, the mass of a particle that *does* remain at rest absolutely, which travels along a fundamental geodesic of the de Sitter sphere and may describe spacetime locally according to the statical SdS solution, is actually zero.

But then in general, it was found that there should be no objective reason to restrict the mass of a particle to positive values, but that the existence of negative mass particles should be expected to be as likely as positive masses according to the symmetry of the local SdS solution. For, since mass was supposed to be directly related to a particle's absolute motion through the cosmic 3-sphere, which is indeed parallelisable, the objective difference between positive and negative masses was interpreted as being due to both the magnitude and direction of their absolute velocity vector.

And by analysing the descriptions of the gravitational field surrounding both positive and negative masses, it was found that the two should appear to be mutually repulsive and invisible. This result was thought to be significant because it could then be interpreted as meaning: if such a universe did contain an initially homogeneous distribution of such particles, then in time these would naturally dissociate into a sort of lattice structure; and, since the expansion of such a universe would be absolute, so that it would not dynamically depend on the material content as it does in FLRW theory, the gravitational interaction between any local cluster thus formed, and an effective surrounding dark matter distribution, would be perceived locally as an additional central force. Thus, it was proposed that this repulsive dark anti-matter should be primarily concentrated in the cosmic voids that are observed in our Universe, and that the apparent void-filament large-scale structure that we observe should therefore be due to the fact that we are only able to see the luminous matter that exists.

Now, the distinction between positive, negative, and zero mass particles is also significant in that it indicates how one might algebraically construct a line-element that would be appropriate to the cosmological description from the perspective of massive particles—viz. as the average rest-frame of one species of particles might be defined as the bundle of null lines which all propagate in the same direction along the de Sitter sphere, with the paths of zero-mass photons subsequently re-defined as null. The general relativistic solution to the line-element thus defined turns out to be the cosmological form of the SdS metric, which components are algebraically similar to those of the local solution, but with the radial coordinate now describing the cosmical evolution of the expanding 3-sphere.

It is important to note, that the essential difference between FLRW cosmology and the present theory may be stated finally with regards to this prescription. For, in deriving the general line-element for the background geometry in the former theory, Robertson made four basic assumptions [27]: (i.) a congruence of geodesics, (ii.) hypersurface orthogonality, (iii.) homogeneity, and (iv.) isotropy. Now, these assumptions *do* apply to the proper coordinate system of comoving photon geodesics in the present theory, in which the metric does indeed have the RW form. However, although the universe is thus essentially supposed to be isotropic and homogeneous, and the diverging bundle of fundamental worldlines for massive particles do satisfy Weyl's principle, due to the assumption of an absolute time, those isotropic and homogeneous cosmic hypersurfaces are not orthogonal—in fact, they're at 45° angles—to the bundle of fundamental worldlines. Incidentally, the cosmic hypersurface is not synchronous in the proper fundamental frame, whereas this is required by Robertson's assumptions (i.) and (ii.), in accordance with the Einsteinian interpretation of the relativity of simultaneity. The other difference, that in the FLRW theory the expansion is described to occur according to the stress-energy of matter in the universe, whereas the expansion occurs as a fundamental property of the vacuum in the present theory, is really only secondary to this point, as, in either case, the restriction of the components of the metric according to Einstein's equation is subsequent to the algebraic statement of the basic line-element.

Now, the facts, that the universe actually is isotropic and homogeneous, and that the fundamental worldlines satisfy Weyl's principle, are actually enough to meet the requirement that this universe should appear isotropic to any massive observer in the fundamental rest-frame, due to the relative co-motion of matter. But also, the factor of expansion should be constrained precisely to the rate at which the perceived cosmic radius (which was defined as the radial coordinate in the cosmological SdS solution) evolves in the proper time of fundamental observers; and, as shown in § 4.3, this is in fact identical to the flat ΛCDM scale-factor of FLRW cosmology.

Therefore, because this was all described to take place only on the expanding half of a de Sitter sphere, beginning from a coordinate singularity at its equator, while the prior contracting half of the sphere should be appropriate to the description of the radial evolution of matter during the final stage of spherically symmetric gravitational collapse, it was proposed in § 4.4 that our Universe, which appears to us just as the expanding universe described in this theory, might have formed through such a local gravitational collapse of matter in a prior one, with that physical process having given rise to the absolute time that needed to be assumed.

References

[1] A. S. Eddington. *The Mathematical Theory of Relativity*. Cambridge University Press, Cambridge, 2nd edition, 1923.

[2] A. S. Eddington. *The expanding universe*. 1933.

[3] Isaac Newton. *Mathematical Principles of Natural Philosophy and The System of the World*. University of California Press, 1966. Translated by Andrew Motte; revised translation by Florian Cajori.

[4] Jonathan Barnes, editor. *The Complete Works of Aristotle: The Revised Oxford Translation*. Princeton University Press, Princeton, New Jersey, 1984.

[5] Albert Einstein and R.W. Lawson. *Relativity: the special and the general theory*. Routledge classics. Routledge, 2001.

[6] H. A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl. *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*. Dover publications, New York, 1952.

[7] C. W. Rietdijk. A rigorous proof of determinism derived from the special theory of relativity. *Philosophy of Science*, 33(4):pp. 341–344, 1966.

[8] Hilary Putnam. Time and physical geometry. *Journal of Philosophy*, 64(8):240–247, 1967.

[9] Howard Stein. On einstein–minkowski space–time. *The Journal of Philosophy*, 65(1):pp. 5–23, 1968.

[10] Howard Stein. On relativity theory and openness of the future. *Philosophy of Science*, 58(2):pp. 147–167, 1991.

[11] Craig Callender. Shedding light on time. *Philosophy of Science*, 67(3):S587–S599, 2000.

[12] Steven Savitt. Being and becoming in modern physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2008 edition, 2008.

[13] A. S. Eddington. *The nature of the physical world*. 1929.

[14] Kip S. Thorne. *Black Holes and Time Warps: Einstein's Outrageous Legacy*. W. W. Norton & Company, Inc., New York, 1994.

[15] D. Finkelstein. Past-Future Asymmetry of the Gravitational Field of a Point Particle. *Physical Review*, 110:965–967, May 1958.

[16] Albrecht Fölsing. *Albert Einstein: A Biography.* Viking, 1997. Translated by Ewald Osers.

[17] Roger Penrose. *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics*. Oxford University Press, Oxford, 1989.

[18] R. Geroch. *General relativity form A to B.* 1978.

[19] Howard Stein. Comments on "the thesis of parmenides". *The Review of Metaphysics*, 22(4):pp. 725–734, 1969.

[20] H. Weyl. Zur allgemeinen Relativitätstheorie. *Physicalische Zeitschrift*, 24:230–232, 1923.

[21] H. Weyl. Republication of: On the general relativity theory. *General Relativity and Gravitation*, 41:1661–1666, June 2009.

[22] H. Weyl. Massenträgheit und Kosmos. *Naturwissenschaften*, 12:197–204, March 1924.

[23] H. Weyl. Redshift and relativistic cosmology. *Philosophical Magazine*, 9:936–943, May 1930.

[24] H. P. Robertson. Relativistic cosmology. *Reviews of Modern Physics*, 5(1):62–90, January 1933.

[25] H. Bondi. *Cosmology.* Cambridge University Press, Cambridge, 1950, 1960.

[26] J. Ehlers. Editorial note to: H. Weyl, On the general relativity theory. *General Relativity and Gravitation*, 41:1655–1660, June 2009.

[27] S. E. Rugh and H. Zinkernagel. Weyl's principle, cosmic time and quantum fundamentalism. *ArXiv e-prints*, June 2010.

[28] George F. R. Ellis. Physics in the real universe: Time and space-time. In Vesselin Petkov, editor, *Relativity and the Dimensionality of the World*, volume 153 of *Fundamental Theories of Physics*, pages 153–161. Springer, 2007.

[29] G. F. Ellis. On the Flow of Time. *ArXiv e-prints*, December 2008.

[30] G. F. R. Ellis and T. Rothman. Time and Spacetime: The Crystallizing Block Universe. *International Journal of Theoretical Physics*, 49:988–1003, May 2010.

[31] H. A. Lorentz. Electromagnetic phenomena in a system moving with any velocity smaller than that of light. *Koninklijke Nederlandse Akademie van Weteschappen Proceedings Series B Physical Sciences*, 6:809–831, 1904.

[32] H. Poincaré and G.B. Halsted. *The foundations of science: Science and hypothesis, The value of science, Science and method*. Science and education. The Science Press, 1913.

[33] John S. Bell. How to teach special relativity. *Progress in Scientific Culture*, 1, 1976.

[34] J. S. Bell. *Speakable and unspeakable in quantum mechanics*. Collected papers on quantum philosophy. Cambridge University Press, 1987.

[35] W.L. Craig and Q. Smith. *Einstein, relativity and absolute simultaneity*. Routledge Studies in Contemporary Philosophy. Routledge, 2008.

[36] J. T. Jebsen. Über die allgemeinen kugelsymmetrischen Lösungen der Einsteinschen Gravitationsgleichungen im Vakuum. *Arkiv för Matematik, Astronomi och Fysik*, 15(18):1–9, 1921.

[37] J. T. Jebsen. On the general spherically symmetric solutions of Einstein's gravitational equations in vacuo. *General Relativity and Gravitation*, 37:2253–2259, December 2005.

[38] G. D. Birkhoff and R. E. Langer. *Relativity and modern physics*. 1923.

[39] S. Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. July 1972.

[40] Paul Arthur Schilpp, editor. *Albert Einstein: Philosopher-Scientist*. Library of living philosophers. Tudor Publishing Company, 2nd edition, 1951.

[41] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman and Company, San Francisco, 1973.

[42] G. Lemaître. Note on de Sitter's universe. *Journal of Mathematics and Physics*, 4(3):188–192, May 1925.

[43] H. P. Robertson. On relativistic cosmology. *Philosophical Magazine*, 5:835–848, May 1928.

[44] H. Bacry and J.-M. Lévy-Leblond. Possible Kinematics. *Journal of Mathematical Physics*, 9:1605–1614, October 1968.

[45] E. Inonu and E. P. Wigner. On the contraction of groups and their representations. *Proceedings of the National Academy of Science*, 39:510–524, June 1953.

[46] Freeman J. Dyson. Missed opportunities. *Bulletin of the American Mathematical Society*, 78:635–652, 1972.

[47] M. Villata. CPT symmetry and antimatter gravity in general relativity. *Europhysics Letters*, 94, April 2011.

[48] M. Villata. Reply to "Comment to a paper of M. Villata on antigravity". *ArXiv e-prints*, September 2011.

[49] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116:1009–1038, September 1998.

[50] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *The Astrophysical Journal*, 517:565–586, June 1999.

[51] A. Einstein. Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, pages 844–847, 1915.

[52] V. M. Slipher. The radial velocity of the Andromeda Nebula. *Lowell Observatory Bulletin*, 2:56–57, 1913.

[53] E. P. Hubble. A spiral nebula as a stellar system: Messier 33. *The Astrophysical Journal*, 63:236–274, May 1926.

[54] E. P. Hubble. Extra-Galactic Nebulae. *The Astrophysical Journal*, 64:321–369, December 1926.

[55] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, 1929.

[56] A. Einstein. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, pages 142–152, 1917.

[57] W. de Sitter. Einstein's theory of gravitation and its astronomical consequences. Third paper. *The Monthly Notices of the Royal Astronomical Society*, 78:3–28, November 1917.

[58] A. Pais. *'Subtle is the Lord ... ': The Science and the Life of Albert Einstein*. 1982.

[59] F. Kottler. Über die physikalischen Grundlagen der Einsteinschen Gravitationstheorie. *Annalen der Physik*, 56:401–462, 1918.

[60] K. Schwarzschild. Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin. Physikalish-Mathematische Klasse*, pages 189–196, 1916.

[61] K. Schwarzschild. On the Gravitational Field of a Mass Point According to Einstein's Theory. *General Relativity and Gravitation*, 35:951–959, May 2003.

[62] G. Lemaître. A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-Galactic Nebulae. *The Monthly Notices of the Royal Astronomical Society*, 91:483–490, March 1931.

[63] G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles A*, 47:49–59, April 1927.

[64] A. Friedman. Über die Krümmung des Raumes. *Zeitschrift für Physik*, 10:377–386, December 1922.

[65] A. Friedman. On the Curvature of Space. *General Relativity and Gravitation*, 31:1991–2000, December 1999.

[66] A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, 21:326–332, December 1924.

[67] A. Friedmann. On the Possibility of a World wih Constant Negative Curvature of Space. *General Relativity and Gravitation*, 31:2001–2008, December 1999.

[68] A. Krasiński and G. F. R. Ellis. Editor's Note: On the Curvature of space and On the Possibility of a World with Constant Negative Curvature of Space. *General Relativity and Gravitation*, 31(12):1985–1990, December 1999.

[69] A. Einstein. Bemerkung zu der Arbeit von A. Friedmann „Über die Krümmung des Raumes". *Zeitschrift für Physik*, 11:326–326, December 1922.

[70] A. Einstein. Notiz zu der Arbeit von A. Friedmann „Über die Krümmung des Raumes". *Zeitschrift für Physik*, 16:228–228, December 1923.

[71] A. S. Eddington. On the instability of Einstein's spherical world. *The Monthly Notices of the Royal Astronomical Society*, 90:668–678, May 1930.

[72] W. de Sitter. On the Distances and Radial Velocities of Extra-Galactic Nebulae, and the Explanation of the Latter by the Relativity Theory of Inertia. *Proceedings of the National Academy of Science*, 16:474–488, July 1930.

[73] G. Lemaître. The Expanding Universe. *The Monthly Notices of the Royal Astronomical Society*, 91:490–501, March 1931.

[74] S. Deser. Introduction to Jebsen's paper. *General Relativity and Gravitation*, 37:2251–2252, December 2005.

[75] G. Lemaître. The Beginning of the World from the Point of View of Quantum Theory. *Nature*, 127:706–706, May 1931.

[76] G. Lemaître. L'Univers en expansion. *Annales de la Societe Scietifique de Bruxelles*, 53:51–85, 1933.

[77] G. A. Lemaître and M. A. H. MacCallum. The Expanding Universe. *General Relativity and Gravitation*, 29:641–680, May 1997.

[78] G. Lemaître. Evolution of the Expanding Universe. *Proceedings of the National Academy of Science*, 20:12–17, January 1934.

[79] A Einstein. Zum kosmologischen problem der allgemeinen relativitätstheorie. *Sitzungsberichte der Preußichen Akademie der Wissenschaften zu Berlin. Physikalisch-Mathematische Klasse*, pages 235–237, 1931.

[80] R. C. Tolman. On the Theoretical Requirements for a Periodic Behaviour of the Universe. *Physical Review*, 38:1758–1771, November 1931.

[81] R. Penrose. Singularities and time-asymmetry. In S. W. Hawking and W. Israel, editors, *General Relativity: An Einstein centenary survey*, pages 581–638, 1979.

[82] O. Heckmann. Über die Metrik des sich ausdehnenden Universums. *Veroeffentlichungen der Universitaets-Sternwarte zu Goettingen*, 2:127–131, June 1931.

[83] O. Heckmann. Die Ausdehnung der Welt in ihrer Abhängigkeit von der Zeit. *Veroeffentlichungen der Universitaets-Sternwarte zu Goettingen*, 2:179–190, February 1932.

[84] H. P. Robertson. On the Foundations of Relativistic Cosmology. *Proceedings of the National Academy of Science*, 15:822–829, November 1929.

[85] H. Weyl. *Raum, Zeit, Materie*. Julius Springer, Berlin, 5th edition, 1923.

[86] R. C. Tolman. On the Possible Line Elements for the Universe. *Proceedings of the National Academy of Science*, 15:297–304, April 1929.

[87] A. Einstein and W. de Sitter. On the Relation between the Expansion and the Mean Density of the Universe. *Proceedings of the National Academy of Science*, 18:213–214, March 1932.

[88] W. de Sitter. *The astronomical aspect of the theory of relativity*. University of California press, Berkeley, 1933.

[89] Albert Einstein. *The Meaning of Relativity*. Princeton University Press, Princeton, N.J., 2nd edition, 1945.

[90] W. de Sitter. On the expanding universe. *Koninklijke Nederlandse Akademie van Weteschappen Proceedings Series B Physical Sciences*, 35:596–607, 1932.

[91] A. G. Walker. On the Formal Comparison of Milne's Kinematical System with the Systems of General Relativity. *The Monthly Notices of the Royal Astronomical Society*, 95:263–269, January 1935.

[92] A. G. Walker. On Riemannian Spaces with Spherical Symmetry about a Line, and the Conditions for Isotropy in General Relativity. *The Quarterly Journal of Mathematics. Oxford Series.*, 6:81–93, 1935.

[93] A. G. Walker. On Milne's Theory of World-Structure. *Proceedings of the London Mathematical Society*, 42:90–127, 1937.

[94] H. P. Robertson. Kinematics and World-Structure. *The Astrophysical Journal*, 82:284–301, November 1935.

[95] H. P. Robertson. Kinematics and World-Structure II. *The Astrophysical Journal*, 83:187–201, April 1936.

[96] H. P. Robertson. Kinematics and World-Structure III. *The Astrophysical Journal*, 83:257–271, May 1936.

[97] S. Chandrasekhar. *Eddington, the most distinguished astrophysicist of his time*. 1983.

[98] S. Weinberg. The Cosmological Constant Problem. *Reviews of Modern Physics*, 61:1–23, January 1989.

[99] T. Padmanabhan. Cosmological constant-the weight of the vacuum. *Physics Reports*, 380:235–320, July 2003.

[100] W. de Sitter. On the expanding universe and the time-scale. *The Monthly Notices of the Royal Astronomical Society*, 93:628–634, June 1933.

[101] W. Pauli. *Theory of Relativity*. Pergamon Press, Princeton, N.J., 1958.

[102] G. Gamow. *My world line: An informal autobiography.* 1970.

[103] S. Weinberg. *Cosmology*. Oxford University Press, Oxford, 2008. A discussion of the cosmological redshift is given in §1.2.

[104] Roger Penrose. *The Road to Reality: A Complete Guide to the Laws of the Universe.* Vintage Books, London, 2005.

[105] R. Penrose. Black holes, quantum theory and cosmology. *Journal of Physics Conference Series*, 174(1):012001–(1–15), June 2009.

[106] A. S. Eddington. The End of the World: from the Standpoint of Mathematical Physics. *Nature*, 127:447–453, March 1931.

[107] L. M. Krauss and G. D. Starkman. Life, the Universe, and Nothing: Life and Death in an Ever-expanding Universe. *The Astrophysical Journal*, 531:22–30, March 2000.

[108] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *The Astrophysical Journal*, 270:365–370, July 1983.

[109] M. Milgrom. A modification of the Newtonian dynamics: Implications for galaxies. *The Astrophysical Journal*, 270:371–383, July 1983.

[110] M. Milgrom. A modification of the Newtonian dynamics: Implications for galaxy systems. *The Astrophysical Journal*, 270:384–389, July 1983.

[111] G. Lemaître. Cosmological Application of Relativity. *Reviews of Modern Physics*, 21:357–366, July 1949.

[112] G. Lemaitre. Clusters of Nebulae in an Expanding Universe. *Monthly Notes of the Astronomical Society of South Africa*, 11:110–117, 1952.

[113] G. Lemaître. Instability in the Expanding Universe and its Astronomical Implications. *Ricerche Astronomiche*, 5:475–488, 1958.

[114] R. C. Tolman. Effect of Inhomogeneity on Cosmological Models. *Proceedings of the National Academy of Science*, 20:169–176, March 1934.

[115] G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results. *The Astrophysical Journal Supplement*, 180:225–245, February 2009.

[116] F. A. E. Pirani. On the influence of the expansion of space on the gravitational field surrounding an isolated body. In *Proceedings of the Cambridge Philosophical Society*, volume 50 of *Proceedings of the Cambridge Philosophical Society*, pages 637–638, October 1954.

[117] Z. Stuchlik. The motion of test particles in black-hole backgrounds with non-zero cosmological constant. *Bulletin of the Astronomical Institutes of Czechoslovakia*, 34:129–149, March 1983.

[118] Z. Stuchlík and S. Hledík. Some properties of the Schwarzschild-de Sitter and Schwarzschild-anti-de Sitter spacetimes. *Physical Review D*, 60(4):044006–(1–15), August 1999.

[119] I. D. Karachentsev, O. G. Kashibadze, D. I. Makarov, and R. B. Tully. The Hubble flow around the Local Group. *The Monthly Notices of the Royal Astronomical Society*, 393:1265–1274, March 2009.

[120] S. E. Rugh and H. Zinkernagel. The quantum vacuum and the cosmological constant problem. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 33(4):663 – 705, 2002. arXiv:hep-th/0012253.

[121] H. Bondi and T. Gold. The Steady-State Theory of the Expanding Universe. *The Monthly Notices of the Royal Astronomical Society*, 108:252–270, 1948.

[122] F. Hoyle. A New Model for the Expanding Universe. *The Monthly Notices of the Royal Astronomical Society*, 108:372–382, 1948.

[123] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. Cosmic Black-Body Radiation. *The Astrophysical Journal*, 142:414–419, July 1965.

[124] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419–421, July 1965.

[125] G. Gamow. Expanding Universe and the Origin of Elements. *Physical Review*, 70:572–573, October 1946.

[126] R. A. Alpher, H. Bethe, and G. Gamow. The Origin of Chemical Elements. *Physical Review*, 73:803–804, April 1948.

[127] G. Gamow. The Evolution of the Universe. *Nature*, 162:680–682, October 1948.

[128] R. A. Alpher and R. Herman. Evolution of the Universe. *Nature*, 162:774–775, November 1948.

[129] R. A. Alpher and R. C. Herman. Theory of the Origin and Relative Abundance Distribution of the Elements. *Reviews of Modern Physics*, 22:153–212, April 1950.

[130] V. Petrosian, E. Salpeter, and P. Szekeres. Quasi-Stellar Objects in Universes with Non-Zero Cosmological Constant. *The Astrophysical Journal*, 147:1222–1226, March 1967.

[131] J. Shklovsky. On the Nature of “Standard” Absorption Spectrum of the Quasi-Stellar Objects. *The Astrophysical Journal Letters*, 150:L1–L3, October 1967.

[132] N. Kardashev. Lemaître’s Universe and Observations. *The Astrophysical Journal Letters*, 150:L135–L139, December 1967.

[133] G. Burbidge. On the Wavelengths of the Absorption Lines in Quasi-Stellar Objects. *The Astrophysical Journal*, 147:851–855, February 1967.

[134] G. R. Burbidge and E. M. Burbidge. Limits to the Distance of the Quasi-Stellar Objects Deduced from Their Absorption Line Spectra. *The Astrophysical Journal Letters*, 148:L107–L113, May 1967.

[135] Ya. B. Zel’Dovich. Cosmological Constant and Elementary Particles. *Soviet Journal of Experimental and Theoretical Physics Letters*, 6:316–317, November 1967.

[136] Ya. B. Zel’dovich. Special Issue: the Cosmological Constant and the Theory of Elementary Particles. *Soviet Physics Uspekhi*, 11:381–393, March 1968.

[137] V. Sahni and A. Krasiński. Republication of: The cosmological constant and the theory of elementary particles (By Ya. B. Zeldovich). *General Relativity and Gravitation*, 40:1557–1591, July 2008.

[138] Ya. B. Zeldovich and I. D. Novikov. *Relativistic astrophysics. Volume 2 - The structure and evolution of the universe*. 1983.

[139] A. H. Guth. *The inflationary universe. The quest for a new theory of cosmic origins.* 1997.

[140] A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23:347–356, January 1981.

[141] Paul J. Steinhardt. The inflation debate. *Scientific American*, 304(4):36–43, April 2011.

[142] I. D. Novikov. *Evolution of the universe.* 1983.

[143] S. Chandrasekhar. The highly collapsed configurations of a stellar mass. *The Monthly Notices of the Royal Astronomical Society*, 91:456–466, March 1931.

[144] S. Chandrasekhar. The highly collapsed configurations of a stellar mass (Second paper). *The Monthly Notices of the Royal Astronomical Society*, 95:207–225, January 1935.

[145] W. M. Wood-Vasey, G. Miknaitis, C. W. Stubbs, S. Jha, A. G. Riess, P. M. Garnavich, R. P. Kirshner, C. Aguilera, A. C. Becker, J. W. Blackman, S. Blondin, P. Challis, A. Clocchiatti, A. Conley, R. Covarrubias, T. M. Davis, A. V. Filippenko, R. J. Foley, A. Garg, M. Hicken, K. Krisciunas, B. Leibundgut, W. Li, T. Matheson, A. Miceli, G. Narayan, G. Pignata, J. L. Prieto, A. Rest, M. E. Salvo, B. P. Schmidt, R. C. Smith, J. Sollerman, J. Spyromilio, J. L. Tonry, N. B. Suntzeff, and A. Zenteno. Observational Constraints on the Nature of Dark Energy: First Cosmological Results from the ESSENCE Supernova Survey. *The Astrophysical Journal*, 666:694–715, 2007.

[146] T. M. Davis, E. Mörtsell, J. Sollerman, A. C. Becker, S. Blondin, P. Challis, A. Clocchiatti, A. V. Filippenko, R. J. Foley, P. M. Garnavich, S. Jha, K. Krisciunas, R. P. Kirshner, B. Leibundgut, W. Li, T. Matheson, G. Miknaitis, G. Pignata, A. Rest, A. G. Riess, B. P. Schmidt, R. C. Smith, J. Spyromilio, C. W. Stubbs, N. B. Suntzeff, J. L. Tonry, W. M. Wood-Vasey, and A. Zenteno. Scrutinizing Exotic Cosmological Models Using ESSENCE Supernova Data Combined with Other Cosmological Probes. *The Astrophysical Journal*, 666:716–725, September 2007.

[147] M. Kowalski, D. Rubin, G. Aldering, R. J. Agostinho, A. Amadon, R. Amanullah, C. Balland, K. Barbary, G. Blanc, P. J. Challis, A. Conley, N. V. Connolly, R. Covarrubias, K. S. Dawson, S. E. Deustua, R. Ellis, S. Fabbro, V. Fadeyev, X. Fan, B. Farris, G. Folatelli, B. L. Frye, G. Garavini, E. L. Gates, L. Germany, G. Goldhaber, B. Goldman, A. Goobar, D. E. Groom, J. Haissinski, D. Hardin, I. Hook, S. Kent, A. G. Kim, R. A. Knop, C. Lidman, E. V. Linder, J. Mendez, J. Meyers, G. J. Miller, M. Moniez, A. M. Mourão, H. Newberg, S. Nobili, P. E. Nugent, R. Pain, O. Perdereau, S. Perlmutter, M. M. Phillips, V. Prasad, R. Quimby, N. Regnault, J. Rich, E. P. Rubenstein, P. Ruiz-Lapuente, F. D. Santos, B. E. Schaefer, R. A. Schommer, R. C. Smith, A. M. Soderberg, A. L. Spadafora, L.-G. Strolger, M. Strovink, N. B. Suntzeff, N. Suzuki, R. C. Thomas, N. A. Walton, L. Wang, W. M. Wood-Vasey, and J. L. Yun. Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets. *The Astrophysical Journal*, 686:749–778, October 2008.

[148] M. Hicken, W. M. Wood-Vasey, S. Blondin, P. Challis, S. Jha, P. L. Kelly, A. Rest, and R. P. Kirshner. Improved Dark Energy Constraints from $\sim$100 New CfA Supernova Type Ia Light Curves. *The Astrophysical Journal*, 700:1097–1140, August 2009.

[149] R. Kessler, A. C. Becker, D. Cinabro, J. Vanderplas, J. A. Frieman, J. Marriner, T. M. Davis, B. Dilday, J. Holtzman, S. W. Jha, H. Lampeitl, M. Sako, M. Smith, C. Zheng, R. C. Nichol, B. Bassett, R. Bender, D. L. Depoy, M. Doi, E. Elson, A. V. Filippenko, R. J. Foley, P. M. Garnavich, U. Hopp, Y. Ihara, W. Ketzeback, W. Kollatschny, K. Konishi, J. L. Marshall, R. J. Mc Millan, G. Miknaitis, T. Morokuma, E. Mörtsell, K. Pan, J. L. Prieto, M. W. Richmond, A. G. Riess, R. Romani, D. P. Schneider, J. Sollerman, N. Takanashi, K. Tokita, K. van der Heyden, J. C. Wheeler, N. Yasuda, and D. York. First-Year Sloan Digital Sky Survey-II Supernova Results: Hubble Diagram and Cosmological Parameters. *The Astrophysical Journal Supplement*, 185:32–84, November 2009.

[150] J. Sollerman, E. Mörtsell, T. M. Davis, M. Blomqvist, B. Bassett, A. C. Becker, D. Cinabro, A. V. Filippenko, R. J. Foley, J. Frieman, P. Garnavich, H. Lampeitl, J. Marriner, R. Miquel, R. C. Nichol, M. W. Richmond, M. Sako, D. P. Schneider, M. Smith, J. T. Vanderplas, and J. C. Wheeler. First-Year Sloan Digital Sky Survey-II (SDSS-II) Supernova Results: Constraints on Nonstandard Cosmological Models. *The Astrophysical Journal*, 703:1374–1385, October 2009.

[151] J. C. Bueno Sanchez, S. Nesseris, and L. Perivolaropoulos. Comparison of recent SnIa datasets. *Journal of Cosmology and Astro-Particle Physics*, 11:29–(1–22), November 2009.

[152] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. *The Astrophysical Journal Supplement*, 192:14–(1–15), February 2011.

[153] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *The Astrophysical Journal Supplement*, 192:18–(1–47), February 2011.

[154] W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *The Monthly Notices of the Royal Astronomical Society*, 401:2148–2168, February 2010.

[155] Pierre-Simon Laplace. *A Philosophical Essay on Probabilities*. Dover Publications, Inc., New York, 6 edition, 1951. Translated by Frederick Wilson Truscott and Frederick Lincoln Emory.

[156] James Clerk Maxwell. *A Treatise on Electricity and Magnetism*. Oxford University Press, London, third edition, 1892.

[157] Sextus Empiricus. *Outlines of Scepticism*. Cambridge University Press, 2000.

[158] Vesselin Petkov, editor. *Relativity and the Dimensionality of the World*, volume 153 of *Fundamental Theories of Physics*. Springer, 2007.

[159] Bertrand Russell. *The ABC of Relativity*. George Allen & Unwin Ltd, 1969.

[160] R. Geroch. Domain of Dependence. *Journal of Mathematical Physics*, 11:437–449, February 1970.

[161] H. G. Alexander, editor. *The Leibniz-Clarke correspondence*. Manchester University Press, 1956.

[162] Don A. Howard. Einstein's philosophy of science. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2010 edition, 2010.

[163] Sextus Empiricus. *Against the Logicians*. Cambridge University Press, 2005. Translated and edited by Richard Bett.

[164] T. M. Robinson. *Heraclitus*. University of Toronto Press, Toronto, 1987.

[165] Jonathan Barnes. *The Presocratic Philosophers*. Routledge & Kegan Paul Ltd., 1982.

[166] C. C. W. Taylor. *The Atomists: Leucuppus and Democritus*. University of Toronto Press, Toronto, 1999.

[167] Diogenes Laërtius. *Lives of Eminent Philosophers*. Harvard University Press, Cambridge, Massachusetts, 1991. In two volumes, with an English translation by R. D. Hicks.

[168] R. D. Hicks. *Epochs of Philosophy: Stoic and Epicurean*. Russell & Russell, Inc., New York, 1962.

[169] John M. Cooper and D. S. Hutchinson, editors. *Plato: Complete Works*. Hackett Publishing Company, Inc., Indianapolis, 1997.

[170] Seneca. *Ad Lucilium Epistulae Morales*. Harvard University Press, Cambridge, Massachusetts, 1917. With an English translation by Richard M. Gummere.

[171] Giorgio de Santillana. *The Origins of Scientific Thought*. The New American Library, New York, 1961.

[172] David Gallow. *Parmenides of Elea*. University of Toronto Press, Toronto, 1984.

[173] Giorgio de Santillana. *The Crime of Galileo.* The University of Chicago Press, Chicago, 1955.

[174] Arthur Koestler. *The Sleepwalkers.* Hutchinson, London, 1959.

[175] Galileo Galilei. *Dialogue Concerning the Two Chief World Systems—Ptolemaic & Copernican.* University of California Press, 2nd edition, 1967. Translated by Stillman Drake, forward by Albert Einstein.

[176] Bertrand Russell. *Fact and Fiction.* George Allen and Unwin, Ltd., London, 1961.

[177] Thomas Kuhn. *The Structure of Scientific Revolutions.* The University of Chicago Press, Chicago, 1962.

[178] A. Einstein. Ernst Mach. *Physikalische Zeitschrift*, 17:101–104, 1916.

[179] A. Einstein. *Ideas and opinions.* Bonanza Books, 1954.

[180] D. Bohm and B. J. Hiley. *The Undivided Universe: An Ontological Interpretation of Quantum Theory.* Routledge, New York, 1993.

[181] J. J. Thomson. *James Clerk Maxwell: a commemoration volume, 1831-1931.* The University press, 1931.

[182] Milič Čapek. *The Philosophical Impact of Contemporary Physics.* D. Van Nostrand Company, Inc., Princeton, 1961.

[183] L. Witten. *Gravitation: an Introduction to current research.* 1962.

[184] R. Arnowitt, S. Deser, and C. W. Misner. Republication of: The dynamics of general relativity. *General Relativity and Gravitation*, 40:1997–2027, September 2008.

[185] John Cowperthwaite Graves. *The Conceptual Foundations of Contemporary Relativity Theory.* The MIT Press, Cambridge, Massachusetts, 1971.

[186] Bertrand Russell. *History of Western Philosophy: and its Connection with Political and Social Circumstances from the Earliest Times to the Present Day.* George Allen and Unwin, Ltd., London, 1946.

[187] St. Augustine. *Confessions.* Harvard University Press, Cambridge, Massachusetts, 1912. With an English translation by William Watts.

[188] J. Ellis McTaggart. The unreality of time. *Mind*, 17(68):457–474, 1908.

[189] A. Einstein. Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 322:891–921, 1905.

[190] Carl Seelig. *Albert Einstein: A Documentary Biography.* Staples Press Limited, London, 1956. Translated by Mervin Savill.

[191] Carl Seelig. *Albert Einstein und die Schweiz.* Europa-Verlag, Zürich, 1952.

[192] Philipp Frank. *Einstein: His Life and Times*. Alfred A. Knopf, New York, 1947. Translated by George Rosen.

[193] Albert Einstein and Leopold Infeld. *The Evolution of Physics: The Growth of Ideas from Early Concepts to Relativity and Quanta*. Simon and Schuster, New York, 1938.

[194] Karl Popper. *Unended Quest: An Intellectual Autobiography*. Routledge Classics, New York, 2002.

[195] H. G. Wells. *The Time Machine: An Invention*. Random House, New York, 1931.

[196] Wikipedia. De Sitter space — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=De_Sitter_spaceoldid=421788618, 2012. [Online; accessed 11-January-2012].

[197] B. G. Schmidt. A new definition of singular points in general relativity. *General Relativity and Gravitation*, 1:269–280, September 1971.

[198] M. D. Kruskal. Maximal Extension of Schwarzschild Metric. *Physical Review*, 119:1743–1745, September 1960.

[199] G. Szekeres. On the singularities of a Riemannian manifold. *Publicationes Mathematicae, Debrecen*, 7:285–300, 1960.

[200] Gy. Szekeres. On the singularities of a Riemannian manifold. *General Relativity and Gravitation*, 34:2001–2016, November 2002.

[201] J. L. Synge. The gravitational field of a particle. *Proceedings of the Royal Irish Academy A*, 53:83–114, 1950.

[202] J. Ehlers and A. Krasiński. Comment on the paper by J. T. Jebsen reprinted in Gen. Rel. Grav. 37, 2253-2259 (2005). *General Relativity and Gravitation*, 38:1329–1330, August 2006.

[203] S. W. Hawking and R. Penrose. The Singularities of Gravitational Collapse and Cosmology. *Royal Society of London Proceedings Series A*, 314:529–548, January 1970.

[204] John Milnor. Analytic proofs of the "hairy ball theorem" and the brouwer fixed point theorem. *The American Mathematical Monthly*, 85(7):pp. 521–524, 1978.

[205] J. Christian. On the Origins of Quantum Correlations. *ArXiv e-prints*, December 2011.

[206] K. Schleich and D. M. Witt. A simple proof of Birkhoff's theorem for cosmological constant. *Journal of Mathematical Physics*, 51(11):112502–(1–9), November 2010.

[207] H. Nariai. On some static solutions of Einstein's gravitational field equations in a spherically symmetric case. *Science Reports of the Tohoku University*, 34(3):160–+, 1950.

[208] H. Nariai. On Some Static Solutions of Einstein's Gravitational Field Equations in a Spherically Symmetric Case. *General Relativity and Gravitation*, 31:951–961, June 1999.

[209] H. Nariai. On a new cosmological solution of Einstein's field equations of gravitation. *Science Reports of the Tohoku University*, 35(1):46–+, 1951.

[210] H. Nariai. On a New Cosmological Solution of Einstein's Field Equations of Gravitation. *General Relativity and Gravitation*, 31:963–971, June 1999.

[211] R. Penrose. On cosmological mass with positive Λ. *General Relativity and Gravitation*, 43:3355–3366, December 2011.

[212] Robert M. Wald. *General relativity*. The University of Chicago Press, Chicago, 1984.

[213] W. Rindler. *Relativity: special, general, and cosmological*. Oxford University Press, Oxford, 2001.

[214] G. W. Gibbons and S. W. Hawking. Cosmological event horizons, thermodynamics, and particle creation. *Physical Review D*, 15:2738–2751, May 1977.

[215] S. L. Bażański and V. Ferrari. Analytic extension of the Schwarzschild-de Sitter metric. *Nuovo Cimento B Serie*, 91:126–142, January 1986.

[216] M. J. Jaklitsch, C. Hellaby, and D. R. Matravers. Particle motion in the spherically symmetric vacuum solution with positive cosmological constant. *General Relativity and Gravitation*, 21:941–951, September 1989.

[217] J. Bičák and J. Podolský. Cosmic no-hair conjecture and black-hole formation: An exact model with gravitational radiation. *Physical Review D*, 52:887–895, July 1995.

[218] A. Corichi and A. Gomberoff. Black holes in de Sitter space: Masses, energies, and entropy bounds. *Physical Review D*, 69(6):064016–(1–15), March 2004.

[219] G. C. McVittie. Laplace's alleged "black hole". *The Observatory*, 98:272–274, December 1978.

[220] T. Padmanabhan. Classical and quantum thermodynamics of horizons in spherically symmetric spacetimes. *Classical and Quantum Gravity*, 19:5387–5408, November 2002.

[221] P. A. M. Dirac. A Theory of Electrons and Protons. *Royal Society of London Proceedings Series A*, 126:360–365, January 1930.

[222] B. F. Schutz. *A First Course in General Relativity*. Cambridge University Press, Cambridge, February 1985.

[223] M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby. *General Relativity*. February 2006.

[224] Francois Schweizer. Colliding and merging galaxies. *Science*, 231(4735):227–234, January 1986.

[225] L. Smolin. *The Life of the Cosmos*. Oxford University Press, 1997.

[226] N. David Mermin. Could Feynman Have Said This? *Physics Today*, 57(5), May 2004.